NUCLEOTIDE SEQUENCE ANALYSIS OF CHLOROPLAST DNA FROM A LIVERWORT, *MARCHANTIA POLYMORPHA* L.

HIDEYA FUKUZAWA

1986

CONTENTS

INTRODUCTION —— 1

CHAPTER I Molecular cloning of promoters functional in Escherichia coli from chloroplast DNA —— 7

CHAPTER II Structure and gene organization of the chloroplast genome —— 19

II-1 Transfer RNA genes —— 37

II-2 Genes for photosynthetic polypeptides —— 44

II-3 Genes for ribosomal proteins and α subunit of RNA polymerase (rpoA) —— 55

II-4 Putative gene ndh3 and unidentified open reading frames —— 61

CHAPTER III Split gene for chloroplast ribosomal protein S12 —— 63

REFERENCES —— 73

SUMMARY —— 86

LIST OF PUBLICATIONS —— 88

ACKNOWLEDGMENT —— 90

ABBREVIATIONS

DNA	deoxyribonucleic acid
E. coli	*Escherichia coli*
IPTG	isopropyl-β-D-thiogalactopyranoside
IR	inverted repeat
LSC	large single copy
M. polymorpha	*Marchantia polymorpha* L.
OD	optical density
ONPG	o-nitrophenyl-β-D-galactopyranoside
ORF	open reading frame
RNA	ribonucleic acid
Rubisco	ribulose-1,5-bisphosphate carboxylase/oxygenase
SD	Shine-Dalgarno
SDS	sodium dodecyl sulfate
SSC	small single copy
Tris	tris(hydroxymethyl)aminomethane
bp	base pairs
kb	kilobase pairs
kd	kilodaltons
mRNA	messenger RNA
rRNA	ribosomal RNA
tRNA	transfer RNA

Gene symbols

atp	genes for subunits of H^+-ATP synthase
infA	gene for initiation factor 1
ndh	ORFs homologous to human mitochondrial NADH dehydrogenase
pet	genes for photoelectron transfer polypeptides
psa	genes for photosystem I chlorophyll *a* apoproteins
psb	genes for photosystem II chlorophyll *a* apoproteins
rbcL	gene for the large subunit of Rubisco
rpl	genes for 50S subunit of ribosomal proteins
rpo	genes for subunits of RNA polymerase
rps	genes for 70S subunit of ribosomal proteins
rrn	genes for ribosomal RNAs
trn	genes for transfer RNAs

INTRODUCTION

The chloroplast is the organelle of plants and green algae that has the photosynthetic apparatus and its own house keeping machinery. Chloroplasts and other plastids contain their own autonomously replicating DNA genome that is composed by double stranded covalently closed circular DNA. In 1960s, DNA molecules were visualized in chloroplasts by electronmicroscopic analysis (Ris and Plaut 1962, Sager and Ishida 1963). In 1976, maize chloroplast DNA was analyzed by restriction enzymes and physically mapped to be a circle of about 140,000 nucleotide pairs (Bedbrook et al. 1976). This included two large inverted repeat regions (IR_A and IR_B) of 22 kb coding for a set of ribosomal RNA genes. These IR regions were separated by large and small single copy regions (LSC and SSC regions). Chloroplast DNA of most plants and green algae has chromosomes of this general structure except for a few legumes (Pisum sativum and Vicia faba) and Euglena gracilis. Pea chloroplast DNA has no large repeat regions and Euglena gracilis chloroplast DNA has tandemly repeated regions coding for ribosomal RNA operons. Among chloroplast DNAs of the inverted repeat type from green plants, one of the smallest size is that of Marchantia polymorpha (121 kb) with 10 kb IR sequences (Ohyama et al. 1983), whereas that of Chlamydomonas reinhardii is the largest (195 kb) with 21 kb IR sequences (Rochaix and Malnoe 1978).

The majority of proteins present in chloroplasts are encoded by nuclear DNA, but the rest are encoded by chloroplast DNA and synthesized by the chloroplast transcription-translation machinery. The nucleotide sequences of many chloroplast genes from various plant species have been determined (Crouse et al. 1985). The first chloroplast gene sequence to be determined for a chloroplast protein was that for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (rbcL) (McIntosh et al. 1980). Since that time, a sizable number of plastid genes for proteins have been sequenced as summarized in Table 1. However, the complete sequence has not been determined as yet for any species of plants.

Table 1. Genes for proteins coded by chloroplast genome.

Gene	Protein product	Plant source	Reference	
rbcL	Ribulose-1,5-bisphosphate carboxylase/oxygenase	Maize	McIntosh et al.	1980
		Spinach	Zurawski et al.	1981
		Tobacco	Sinozaki & Sugiura	1982
		Anabaena	Curtis & Haselkorn	1983
		Chlamydomonas reinhardii	Dron et al.	1983
		Synechcoccus	Reichelt & Delaney	1983
		Anacyctis nidulans	Shinozaki et al.	1983
		Rhodospirillum rubum	Nargang et al.	1984
		Barley	Zurawski et al.	1984
		Euglena gracilis	Gingrich & Hallick	1985
psaA	PS I, P700 apoprotein	Maize	Fisch et al.	1985
		Spinach	Kirsch et al.	1986
psaB	PS I, P700 apoprotein	Maize	Fisch et al.	1985
		Spinach	Kirsch et al.	1986
psbA	PS II, "32 kd" protein	Nicotiana debneyi	Zurawski et al.	1982
		Spinach	Zurawski et al.	1982
		Amaranthus hybridus	Hirshberg & McIntosh	1983
		Soybean	Spielmann & Stutz	1983
		Anabaena	Curtis & Haselkorn	1984
		Chlamydomonas reinhardii	Erickson et al.	1984
		Pea	Oishi et al.	1984
		Solanum nigrum	Goloubinoff et al.	1984
		Euglena gracilis	Keller & Stutz	1984
		Nicotiana tabacum	Sugita & sugiura	1984
		Mustard	Link & Langridge	1984
psbB	PS II, P680 "51 kd" protein	Spinach	Morris & Herrmann	1984
psbC	PS II, "44 kd" protein	Spinach	Alt et al.	1984
		Spinach	Holschuh et al.	1984
psbD	PS II, "D2 polypeptide"	Chlamydomonas reinhardii	Rochaix et al.	1984
		Pea	Rasmussen et al.	1984
		Spinach	Alt et al.	1984
		Spinach	Holschuh et al.	1984
psbE	PS II, cytochrome b-559	Spinach	Herrmann et al.	1984
		Oenothera hookeri	Carrillo et al.	1986
		Nicotiana tabacum	Carrillo et al.	1986
		Wheat	Hird et al.	1986
psbF	PS II, cytochrome b-559 (URF39)	Spinach	Herrmann et al.	1984
		Oenothera hookeri	Carrillo et al.	1986
		Nicotiana tabacum	Carrillo et al.	1986
		Wheat	Hird et al.	1986
psbG	PS II, G-protein	Maize	Steinmetz et al.	1986

Table 1 (continued).

Gene	Protein product	Plant source	Reference	
petA	Cytochrome f preprotein	Pea	Willey et al.	1984
		Spinach	Alt et al.	1984
		wheat	Willey et al.	1984
		Oenothera hookeri	Tyagi & Herrmann	1986
petB	Cytochrome b6	Spinach	Heinemeyer et al.	1984
petD	Cytochrome b6/f complex, subunit 4	Pea	Phillips & Gray	1984
		Spinach	Heinemeyer et al.	1984
atpA	H^+-ATP synthase, alpha subunit	Tobacco	Deno et al.	1983
atpB	H^+-ATP synthase, behta subunit	Maize	Krebbers et al.	1982
		Spinach	Zurawski et al.	1982
		Tobacco	Shinozaki & Sugiura	1983
		Barley	Zurawski & Clegg	1984
atpE	H^+-ATP synthase, epsilon subunit	Maize	Krebbers et al.	1982
		Spinach	Zurawski et al.	1982
		Tobacco	Shinozaki et al.	1983
		Barley	Zurawski & Clegg	1984
atpF	H^+-ATP synthase, subunit I	Wheat	Bird et al.	1985
		Tobacco	Shinozaki et al.	1986
atpH	H^+-ATP synthase, subunit III	Wheat	Howe et al.	1982
		Spinach	Alt et al.	1983
		Tobacco	Deno et al.	1984
atpI	H^+-ATP synthase, subunit IV	Pea	Cozens et al.	1986
rpl2	50S ribosomal protein 2	Nicotiana debneyi	Zurawski et al.	1984
		Spinach	Zurawski et al.	1984
rpl16	50S ribosomal protein 16	Spirodela oligorhiza	Posno et al.	1986
rps4	30S ribosomal protein 4	Maize	Subramanian et al.	1983
rps7	30S ribosomal protein 7	Euglena gracilis	Montandon & Stutz	1984
rps11	30S ribosomal protein 11	Spinach	Muller et al.	1986
rps12	30S ribosomal protein 12	Euglena gracilis	Montandon & Stutz	1984
rps14	30S ribosomal protein 14	Marchantia polymorpha	Umesono et al.	1984
		Spinach	Kirsch et al.	1986
rps16	30S ribosomal protein 16	Tobacco	Shinozaki et al.	1986
rps19	30S ribosomal protein 19	Tobacco	Sugita & Sugiura	1983
		Nicotiana debneyi	Zurawski et al.	1984
		Spinach	Zurawski et al.	1984
infA	initiation factor 1	Spinach	Muller et al.	1986
tufA	Elongation factor Tu	Euglena gracilis	Montandon & Stutz	1983
rpoA	RNA polymerase α subunit	Spinach	Muller et al.	1986
rpoB	RNA polymerase β subunit	Tobacco	Ohme et al.	1986

The chloroplast DNAs from common bean, soybean, the fern Osmunda and the algae Chlamydomonas reinhardii have been shown to exist as a 50:50 mixture of two genetically identical but physically distinct molecules called isomer differing only for in the relative orientation of their single copy regions (Palmer 1983). In the case of M. polymorpha, isomers have been identified by fragmentation using PstI, and southern hybridization analysis of the chloroplast DNA with cloned DNA fragments as shown in Fig. 1. ^{32}P-Labeled DNA fragments involved in either single copy region (probe 1) or inverted repeat (probe 2) hybridized to the chloroplast DNA segments generated by BamHI or PstI restriction enzyme (Fig. 2). Probe 1 containing a DNA segment of an inverted repeat hybridized with the four largest fragments (P1, P1', P2 and P2') produced by PstI. On the other hand, probe 2 containing a DNA segment of a single copy region only hybridized with two fragments (P1 and P1'). So in the M. polymorpha chloroplasts two types of DNA molecules exist in an equal amount. It is interesting to know whether the events on DNA replication have any relationship to the formation of isomers. But the mechanisms of flip-flop recombination have not been clarified yet.

The chloroplast biogenesis depends on the coordinative gene expression of both the chloroplast genome and the nuclear genome. But the interaction between chloroplast and nuclear genetic systems has not been clarified. In this study, the author has intended to reveal the mechanisms of gene expression of the chloroplast genome on the basis of the molecular biological methods as follows.
Chapter I; Chloroplast transcriptional promoters were cloned and characterized using the E. coli gene fusion method. Chapter II; The primary structure of the M. polymorpha chloroplast DNA was determined and genes for transfer RNAs, photosynthetic polypeptides, ribosomal proteins, and α subunit of RNA polymerase were identified. Chapter III; An unique system for gene expression that may be called trans-splicing was found in the biogenesis of ribosomal protein S12.

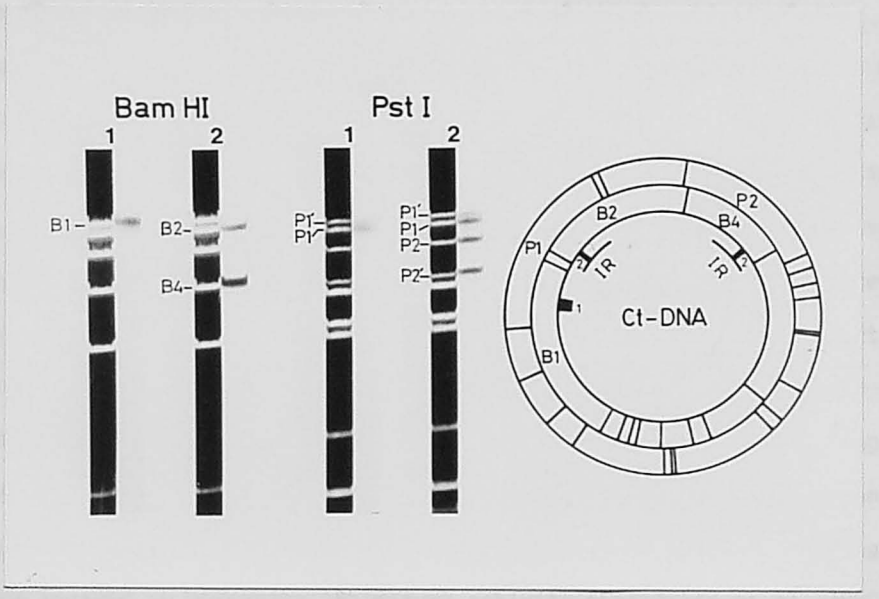
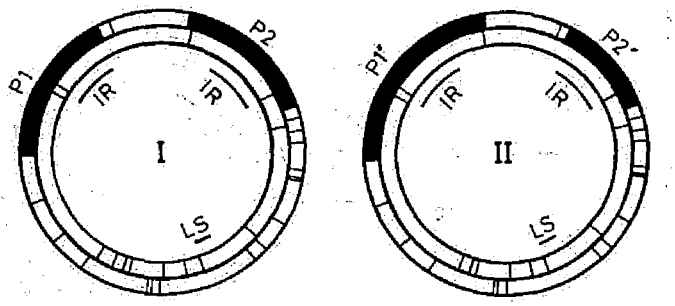


Figure 1. Southern hybridization of the probes containing DNA segments of an inverted repeat region and a single copy region. Chloroplast DNA fragments digested by BamHI and PstI restriction enzymes were electrophoresed in 0.6% agarose gels (each left lane), transferred to nylon membrane filters and hybridized with ^{32}P-labeled probes. Filters were hybridized with the probe 1 containing the Bg16 segment of a single-copy region (black box in the right figure and Lane 1) and probe 2 containing the Bg21 fragment of an inverted repeat (Lane 2). Autoradiograms are shown right side of each electrophoretic pattern.



Isomers in chloroplast DNA
from M. polymorpha

Figure 2. Isomers and their restriction maps of the chloroplast DNA. Two kinds of molecules (form I and II) of the chloroplast DNA are shown. IR and LS indicate inverted repeat sequences containing ribosomal RNA genes, the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase, respectively.

CHAPTER I Molecular cloning of promoters functional in Escherichia coli from chloroplast DNA

Chloroplasts have their own transcriptional and translational systems including unique transfer RNAs, ribosomal RNAs, RNA polymerase and ribosomal proteins different from those of the cytoplasm. Determination of the locations of transcriptional and translational regulatory regions on the chloroplast genome would lead to a basic understanding of the mechanisms of gene expression in chloroplasts. It has been suggested that mechanisms of gene expression in chloroplasts are similar to those in E. coli (Kidd and Bogorad 1979, Whitfeld and Bottomley 1983, Gruissem et al. 1983, Gruissem and Zurawski 1985). If this is the case, E. coli plasmid vectors can be used to clone chloroplast DNA fragments containing transcriptional start signals which can function in E. coli. Casadaban et al. constructed a plasmid in which β-galactosidase activity was expressed by insertion of exogenous DNA fragments having promoter sequences, ribosomal binding sites, and translation start codons in the right frame, and showed that the strength of promoters and ribosome binding activity can be assayed by measuring the enzyme activity (Casadaban et al. 1980).

In this chapter, the author shows an evidence that some of the chloroplast promoters can be expressed in E. coli by using gene fusion techniques with the β-galactosidase gene. The nucleotide sequence of the DNA fragment with the highest enzyme activity in E. coli was determined and the location of the promoter was mapped on the chloroplast genome. To find out whether the chloroplast DNA fragment in fact initiates transcription in E. coli as well as in chloroplasts, S1 nuclease mapping experiments were done. The mechanisms of gene expression in chloroplasts are discussed.

MATERIALS AND METHODS

DNA isolation, restriction endonuclease digestion, and hybridization

Chloroplast DNA was isolated from liverwort, M. polymorpha suspension culture by the method describes previously (Ohyama et al. 1982) with some modifications.

Chloroplasts were isolated from cultured cells by using blender containing sea sand (20-25 mesh). Restriction digestion, hybridization and plasmid DNA isolation were carried out by the procedure described by Maniatis et al.(1982).

Cloning of chloroplast DNA fragments

Chloroplast DNA was cleaved by EcoRI, Sau3A, BglII, SmaI and BamHI restriction endonucleases. Restricted chloroplast DNA fragments were ligated into dephosphorylated EcoRI, BamHI and SmaI sites of plasmid pMC1403 (Casadaban et al. 1980). The ligated DNA was used to transform competent cells of E. coli strain MC1061 (araD139, Δ(ara, leu)7697, ΔlacX74, galU⁻, galK⁻, hsr⁻, hsm⁻, strA) (Casadaban and cohen 1980). Transformed cells were selected on lactose MacConkey agar plates containing 100 μg/mg of ampicillin. Red colonies using lactose (lac^+) and resistant to ampicillin were picked up and cultured in L-broth in the presence of ampicillin.

β-Galactosidase assay

E. coli cells harboring recombinant plasmids were grown to 2-5 x 10^8 cells/ml and β-galactosidase activities were assayed as described by Platt et al.(1972).

DNA sequencing

The nucleotide sequence was determined by the dideoxy chain termination method (Sanger et al. 1977) using M13 phages mp10 and mp11 (Messing 1983). For size markers in S1 nuclease mapping, a HincII-EcoRI fragment (see Fig. 3) was labeled with [γ^{32}P]-ATP (3000 Ci/mmol, Amersham) at the 5' terminus by T4 polynucleotide kinase after alkali phosphatase treatment, and sequenced by the method of Maxam and Gilbert (1977).

RNA isolation and S1 nuclease mapping

Chloroplast RNA was extracted from chloroplast pellets by phenol chloroform

extraction described by Yamano et al.(1984). E. coli total RNA was extracted by the SDS-hot phenol method and S1 mapping was performed by the procedure described by Aiba et al.(1981) using 5' end labeled EcoRI-Sau3A fragment (10^{5-6} cpm/μg) and 100 μg of RNA.

RESULTS

Cloning and characterization of chloroplast DNA fragments having promoter function

Eleven recombinants were selected on lactose MacConkey agar plates containing ampicillin as red colonies showing β-galactosidase production in E. coli. The β-galactosidase assays of recombinants showed a variety of levels of the activities summarized in Table 1. The recombinants can be grouped in three categories by comparing with IPTG induced β-galactosidase activity in E. coli wild strain W3110: 1, a group with a slightly higher level of the enzyme activity such as recombinants harboring plasmids pMP901-903; 2, a group with a rather low level of the activity such as recombinants carrying plasmids pMP921, pMP954-956; and 3, a group with a very low level of the activity such as recombinants carrying plasmids pMP910-912, pMP953.

Restriction analysis of recombinant plasmids (pMP901-903) with high levels β-galactosidase activity showed that they had the same 5.1 kb DNA fragment, which corresponded to Ec8 fragment (see Fig. 3). To confirm the inserted DNA fragment in the plasmid pMC1403 was generated from M. polymorpha chloroplast DNA and to find its site on the chloroplast physical maps, plasmid pMP903 was labeled with [α^{32}P]-dCTP using the nick translation reaction and hybridized with chloroplast DNA fragments on nitrocellulose filter (Fig. 1). ^{32}P-Labeled pMP903 hybridized to the chloroplast DNA fragments Ba2, Ba4, Ba6 and Ba11 (Fig. 1B). The 1.1 kb BamHI fragment derived from the 5.1 kb EcoRI fragment with an additional 10 bp of pMC1403 was recloned into pMC1403 and the recombinant plasmid was named pMP904 (Fig. 2). Southern hybridization with chloroplast DNA fragments showed that the inserted DNA fragment

Table 1. β-Galactosidase activity and copy numbers in clones containing recombinant plasmids.

Host strains	Plasmids	β-Galactosidase activity (units)[a]	Relative amounts of plasmids[b]
W3110 (−IPTG)	—	<11	—
W3110 (+IPTG)	—	344±28	—
MC1061	—	<5	—
MC1061	pMC1403	<6	1.0
MC1061	pMP901	361±20	0.9±0.1
	pMP902	328±27	0.9±0.1
	pMP903	360±23	0.9±0.1
	pMP904[c]	408±39	1.0±0.1
	pMP905[c]	365±20	0.9±0.1
	pMP906[c]	308±33	1.0±0.1
MC1061	pMP954[d]	200±14	1.0±0.1
	pMP955[d]	267±11	0.9±0.1
	pMP956[d]	233±24	0.8±0.2
	pMP921	253±20	1.1±0.1
MC1061	pMP953[e]	49± 7	0.7±0.2
	pMP910	17± 3	0.9±0.1
	pMP911	73± 8	0.9±0.1
	pMP912	56±11	1.0±0.1

a) Numbers are averages of three experiments and expressed as units per mg of total protein (Platt et al. 1972).

b) Plasmid copy numbers were measured by the method of Projan et al. (1983) and expressed as the number relative to pMC1403 in E. coli strain MC1061 as 1.0.

c) Derived from pMP903 (see Figure 2)

d) Containing rbcL gene.

e) Containing β subunit gene of H^+-ATP synthase.

(1.1 kb) of pMP904 hybridized to Ec4 (located on the other IR region), Ec8, Ba6 and Ba11 (see Fig. 1C). These results and no cleavage site of EcoRI on the Ba11 fragment indicate that the cloned DNA fragment originates from the Ba6 fragment and is located on the large single copy region close to the inverted repeat (IR) (Fig. 3). As the clone harboring plasmid pMP904 still expressed the high level of the enzyme activity, the internal HincII-EcoRI fragment was further subcloned into pMC1403 (Fig.2). The recombinant plasmid, named pMP905, kept its high level of β-galactosidase production. However, the recombinant plasmid pMP906, which was constructed by deletion of HincII-AluI fragment (position at 3 to 348) from pMP905 (see Fig. 4), gave slightly less activity than plasmids pMP904 and pMP905 (Table 1).

Nucleotide sequence of a promoter and its downstream region in chloroplast DNA

As the HincII-EcoRI chloroplast DNA fragment inserted pMP905 appeared to have a transcriptional and translational start signals, the nucleotide sequence between two HincII sites was determined by the strategy shown in Fig. 3. The nucleotide sequence of the 1192 bp chloroplast DNA fragment and junction region between the chloroplast DNA fragment and the lac'Z gene of pMP905 is shown in Fig. 4A and 4B, respectively.

S1 nuclease mapping

The position in the sequence corresponding to the 5' end of the mRNA was determined by an S1 nuclease mapping procedure. The 533 bp HincII-EcoRI fragment (position at 4 to 536 in Fig. 4A) was labeled at the 5' ends with [γ^{32}P]-ATP and digested by Sau3A restriction enzyme, generating two fragments, 66 nucleotides long and 467 nucleotide long. The ^{32}P-labeled 467 bp fragment was hybridized with either E. coli RNA or chloroplast RNA at 37°C. Hybridized DNA-RNA molecules were digested by S1 nuclease and the length of the S1 nuclease-resistant DNA fragment was measured by electrophoresis with a Maxam-Gilbert sequence ladder generated from the 467 bp fragment as shown in Fig. 5. The 5' end of the mRNA from both E. coli and

Figure 1. Southern hybridization of chloroplast DNA fragments with plasmids pMP903 (B) and pMP904 (C). Panel A shows agarose gel electrophoresis stained by ethidium bromide. Lanes 1, 2, 3, 4 and 5 correspond to the electrophoretic patterns of digests by EcoRI, EcoRI/BamHI, BamHI, BamHI/BglII and BglII restriction enzymes, respectively.

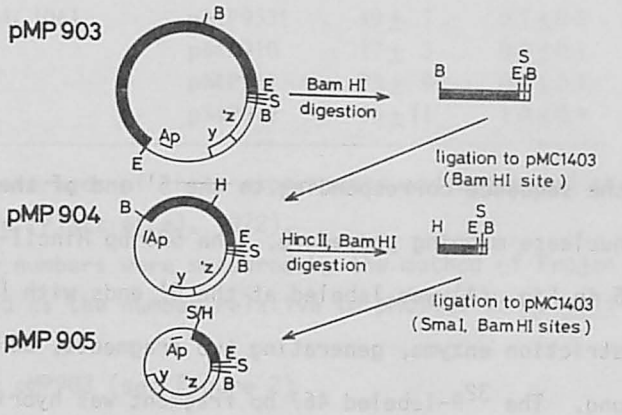


Figure 2. Construction of plasmids pMP904 and pMP905 from pMP903 containing promoter regions. Heavy lines indicate chloroplast DNA segments. The structural gene of β-galactosidase is shown as 'Z.

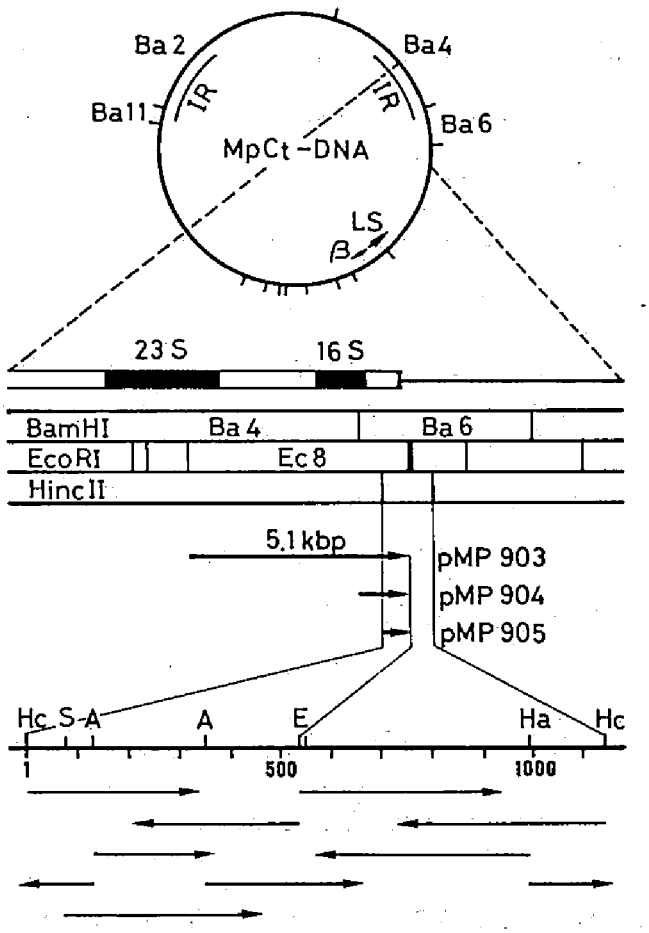


Figure 3. Location of the DNA fragments having promoter function on the physical map of chloroplast DNA, and the sequencing strategy. IR indicates inverted repeat region containing ribosomal RNA operons including 16S and 23S rRNA genes. LS and β indicate the genes for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase and for the β subunit of H^+-ATP synthase, respectively. A, E, Ha, Hc and S indicate cleavage sites of the restriction enzymes AluI, EcoRI, HaeIII, HincII and Sau3A, respectively. Bottom arrows indicate sequencing strategies using the M13 phages mp10 and mp11.

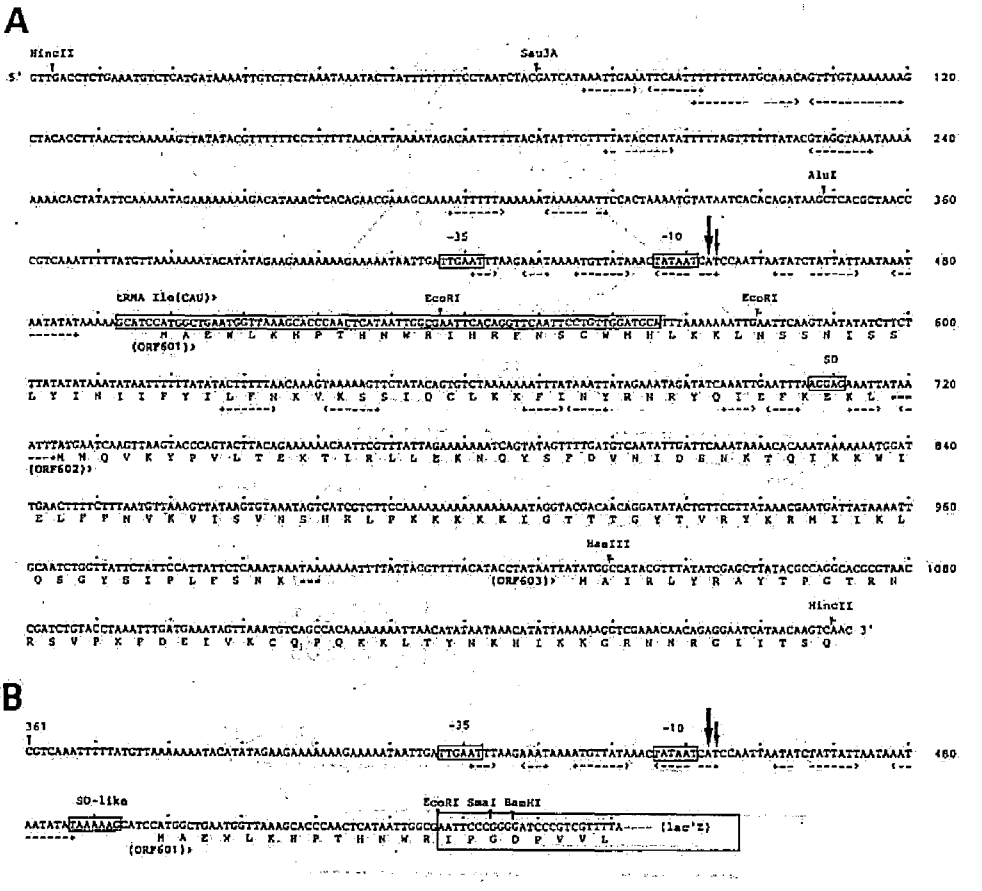


Figure 4. Nucleotide sequence of a DNA fragment having promoter function derived from liverwort chloroplast DNA. Nucleotide sequence of the 1192 bp HincII fragment (Fig. 3) is shown in (A). Possible stem and loop structures are shown by broken lines with arrows. Bold vertical arrows indicate the start sites of mRNA transcription in E. coli as well as in chloroplasts. The regions of "-35", "-10", Shine-Dalgarno sequence (SD), and genes for $tRNA^{Ile}(CAU)$ and β-galactosidase are boxed. Deduced amino acid sequences of ORF601, 602 and 603 are shown under the nucleotide sequences by single letter symbols. Double underlines indicate termination codons. N-Terminal region of the fused protein between ORF601 and β-galactosidase coded by pMP903-906 is shown in (B).

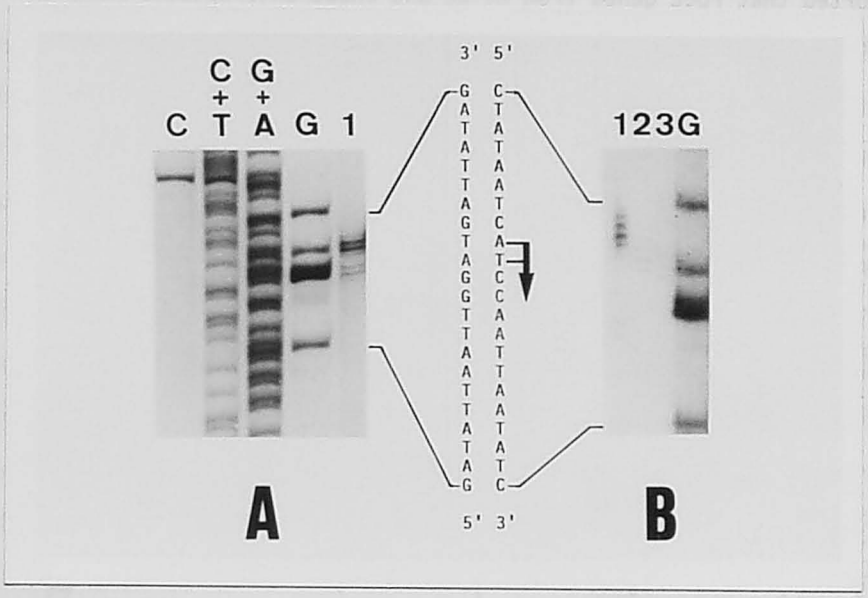


Figure 5. S1 nuclease mapping by E. coli RNA (A) and chloroplast RNA (B) of the promoter region. The S1 nuclease protected DNA fragments (lanes 1, 2 and 3) were electrophoresed in parallel with the Maxam-Gilbert sequence ladders. Lane 1 (A), and lanes 1, 2 and 3 (B) correspond to the concentration of S1 nucleases, 5000, 50, 500 and 5000 units per reaction, respectively. S1 mapping in lane 1 (B) should read to be A, because of the smiling pattern of the gel electrophoresis. An arrow indicates the direction of the transcription.

chloroplast was thus mapped on the sequence to be 45-46 nucleotides upstream from the ATG translational start codon of ORF601 (Fig. 4).

DISCUSSION

From M. polymorpha chloroplast DNA, DNA fragments were cloned functional in E. coli transcription and translation system. Several chloroplast genes have been cloned into E. coli plasmids. With the rbcL gene, its expression was observed in an in vitro coupled transcriptional translational system derived from E. coli. Gatenby

et al. reported that rbcL genes from maize and wheat chloroplasts were expressed in E. coli (1981). Kong et al. reported the cloning of promoter-containing restriction fragments from Nicotiana chloroplast DNA and location of the fragments on the chloroplast genome (1984).

In this study 11 recombinants were obtained which were selected on lactose MacConkey plates as red colonies. These clones, however, varied in their enzyme activity (Table 1). As plasmid copy numbers were not much different in each clone, those variation may reflect the efficiency of the transcriptional start signals and ribosome binding activities in the chloroplasts. For instance, a recombinant carrying plasmid pMP954-956 (containing the rbcL promoter) had a rather high level of the enzyme activity as expected. A recombinant harboring plasmid pMP953 (containing the promoter region of the β subunit gene of H^+-ATP synthase) showed quite a low level of the activity. These results coincide with the fact that the mRNA synthesis of the β subunit gene is considerably lower than that of rbcL gene (Shinozaki et al. 1983). Therefore, the efficiency of E. coli transcriptional system may reflect that of the transcription in chloroplasts.

The plasmid pMP905, carrying a promoter region of an unidentified open reading frame named ORF601, gave the highest level of enzyme activity in E. coli. Analysis of the nucleotide sequence of the promoter and its downstream region revealed that a translational initiation codon (ATG) of ORF601 was found 38 bp upstream from the EcoRI site and its open reading frame was fused to the lac'Z gene in the right frame (see Fig. 4B). And 12 bp upstream from ATG codon, a sequence (TAAaaAG) partially complementary to the 3' end of E. coli 16S rRNA (Shine-Dalgarno (SD) sequence) (Shine and Dalgarno 1974) was found. There was a typical sequence for the transcriptional promoter signal (TATAAT), called the Pribnow-box (Pribnow 1975) or "-10" region at 53 bp upstream from the ATG codon. There was an unique sequence (aTTGAat) at 82 bp upstream, called the "-35" region which is thought to be a RNA polymerase recognition site in E. coli (Takanami et al. 1976). In addition, three possible stem-loop structures can be formed between the "-35" region and SD-like

sequence as indicated by underlining with arrows in Fig. 4A. But at the position 493-566 a tRNA gene, whose anticodon was CAU, was identified by forming typical secondary structure as shown in Fig. 6. This tRNA gene has 94.6% homology (79/84) with spinach chloroplast isoleucine tRNA gene (Francis and Dudock 1982) located in IR regions. So this tRNA gene was confirmed to code isoleucine tRNA in M. polymorpha chloroplasts. This highly active promoter may be for $tRNA^{Ile}$(C*AU) and for the downstream ORFs in the LSC region. Three open reading frames were identified downstream from the promoter region, and designated ORF601, ORF602 and ORF603. Their possible gene products were estimated to be 73, 91 and >52 amino acid residues, respectively. A typical SD sequence (AGGAG) was found seventeen bp upstream from the ATG codon of ORF602, and a stem structure can be formed between the SD sequence and ATG codon as reported in the Chlamydomonas rbcL gene (Dron et al. 1983). This stem structure may have a role to let the SD sequence close to the ATG codon. ORF602 and ORF603 were identified to be putative genes for proteins corresponding to E. coli ribosomal protein L23 and L2, respectively. Detail data are presented in chapter II.

Nucleotide sequence analysis revealed that the organization of this chloroplast promoter was similar to that of E. coli promoters. Furthermore, these results of S1 mapping using in vivo transcripts from both E. coli and chloroplasts showed that the transcription starts at almost the same position downstream from the promoter region. So the gene fusion method described here could be a powerful technique to clone and characterize the promoters on the chloroplast genome.

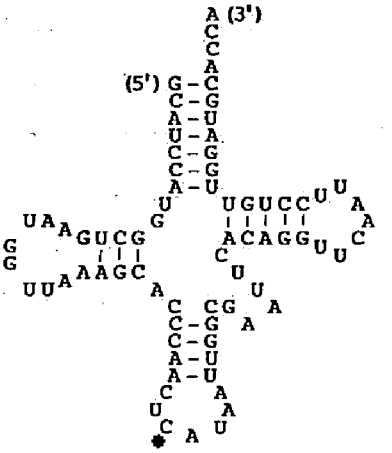


Figure 6. Secondary structure of M. polymorpha chloroplast tRNA^{Ile}(C*AU) deduced from the DNA sequence. The nucleotides CCA at the 3' terminus are not encoded by the chloroplast genome. C* is putative hypermodified base.

Chapter II Structure and gene organization of the chloroplast genome

To understand the genetic system in the chloroplast, the nucleotide sequence of the liverwort, M. polymorpha chloroplast DNA was determined. The M. polymorpha chloroplast DNA has been physically mapped previously (Ohyama et al. 1983). The gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase has been mapped on the chloroplast genome by heterologous hybridization with tobacco rbcL gene. The genes for ribosomal RNAs; 23S, 16S, 5S and 4.5S have been also localized in the inverted repeat regions (Ohyama et al. 1983, Yamano et al. 1984, Yamano et al. 1985). The overall gene organization deduced from the complete nucleotide sequence is described by Ohyama et al. (1986). In this study, properties and characterization of genes on the LSC region (from psbG to 16S rRNA gene; 30,600 bp) deduced from the nucleotide sequence are presented and discussed. The region analyzed in this study is shown as a black box with physical maps in Fig. 1. In the region described here, putative genes for seven tRNAs, ten photosynthetic polypeptides, thirteen ribosomal proteins and α subunit of RNA polymerase were identified. In addition, an open reading frame (ORF) was found to show significant amino acid sequence homology to a subunit of NADH dehydrogenase in human mitochondria.

MATERIALS AND METHODS

Chloroplast DNA was isolated from cell suspension culture of M. polymorpha as described previously (Ohyama et al. 1982). Chloroplast DNA was cloned into E. coli plasmid vectors; pBR322, pKC7, pUC13, pUC18 and pUC19. Recombinant plasmids used for the nucleotide sequencing were summarized in Table 1. The locations of chloroplast DNA fragments used in sequence determination are shown in Fig. 2. Each plasmid was sonicated (Deininger 1983) by TOMY handy sonicator and randomly cloned into SmaI or HincII sites of phage mp18 and mp19 (Perron et al. 1981). Recombinant phages containing chloroplast DNA fragments were screened by dot-hybridization with the chloroplast DNA (Hu and Messing 1982). Obtained shotgun libraries were used for

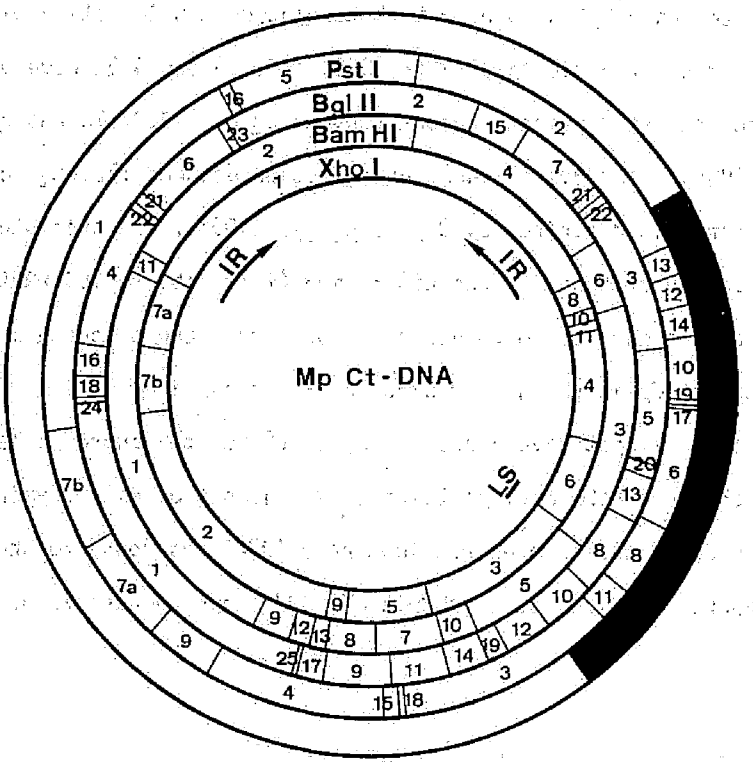


Figure 1. Restriction map of the M. polymorpha chloroplast DNA. Narrow lines with arrow heads inside the circular map indicate inverted repeat regions. LS indicates the site of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (rbcL). The sequenced region in this study is shown as a black box.

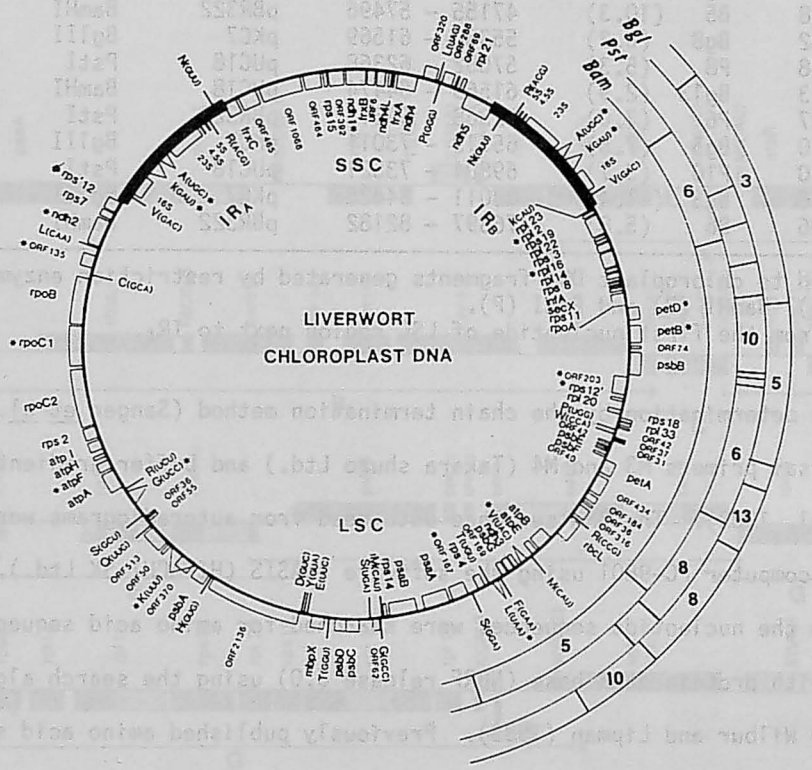


Figure 2. Gene organization of the chloroplast genome from a liverwort, M. polymorpha and sequenced restriction fragments. Thick lines indicate the inverted repeats (IR_A and IR_B). SSC and LSC indicate the small single-copy region and large single-copy region, respectively. Genes shown outside the map are transcribed anticlockwise, and those inside are transcribed clockwise. The tRNA genes are identified by the one-letter amino acid code with their anticodons given in parentheses. Asterisks indicate genes having introns in their sequences (Ohyama et al. 1986). Restriction fragments used in sequence determination are shown outside the genetic map.

Table 1. Recombinant plasmids used for the nucleotide sequence determination.

Plasmid	Fragment[a] (kb)	Position[b]	Vector	Cloning site
pMP593	Bg10 (3.8)	51611 - 55380	pKC7	BglII
pMP228	B5 (10.3)	47155 - 57496	pBR322	BamHI
pMP452	Bg8 (6.2)	55380 - 61569	pKC7	BglII
pMP708	P8 (5.3)	57082 - 62368	pUC18	PstI
pMP713	Bg13 (2.9)	61569 - 64474	pUC18	BamHI
pMP727	P6 (7.0)	62368 - 69315	pBR322	PstI
pMP310	Bg5 (7.5)	65513 - 73011	pKC7	BglII
pMP710	P10 (3.7)	69804 - 73521	pUC18	PstI
pMP376	Bg3 (11.4)	73011 - 84425	pKC7	BglII
pMP206	B6 (5.8)	76397 - 82182	pBR322	BamHI

a) correspond to chloroplast DNA fragments generated by restriction enzymes; BglII (Bg), BamHI (B) and PstI (P).
b) counted from the first nucleotide of LSC region next to IR_A.

DNA sequence determination by the chain termination method (Sanger et al. 1977) using universal primers M3 and M4 (Takara shuzo Ltd.) and buffer gradient gels (Biggin et al. 1983). The DNA sequence data read from autoradiograms were handled by personal computer PC-9801 using the software DNASIS (HITACHI SK Ltd.). ORFs deduced from the nucleotide sequences were searched for amino acid sequence homologies with protein data base (NBRF release 6.0) using the search algorithm described by Wilbur and Lipman (1983). Previously published amino acid sequences of polypeptides in other plant species were also used to homology search.

RESULTS

Gene organization in the LSC region (psbG-16S rRNA gene)

The nucleotide sequence (30,600 bp) from the BglII site (position 51611) to BamHI site (82188) covering the junction (J_{LB}) between LSC and IR_B regions was determined. A computer search of the DNA sequence led to the identification of ORFs that begin with the translation initiation codon AUG and end at an either termination codon (UAA, UAG, and UGA). The gene organization deduced from the nucleotide sequence are shown schematically in Fig. 3. The LSC region sequenced in this study was divided into seven blocks (Fig. 3 A-G) depending on the directions of the translational orientation (shown by horizontal arrows in Fig. 3). The

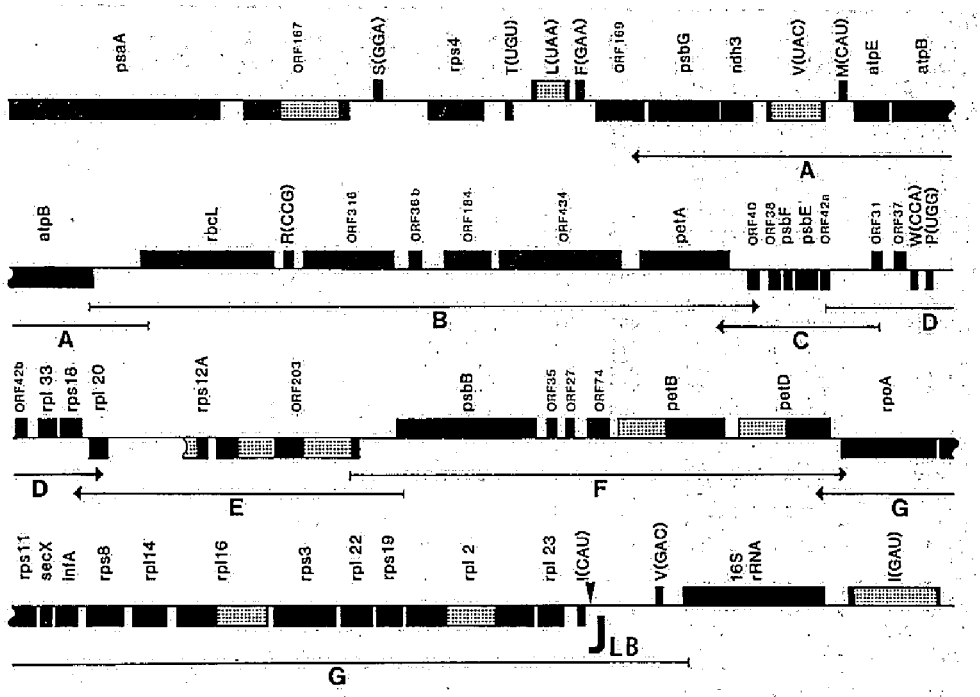


Figure 3. Detail gene organization of the region sequenced in this study. The coding regions of genes are indicated by bold lines. Introns (intervening sequences) are shown as hatched box. Genes shown on lines are transcribed to the right side, and those under lines are transcribed to the left side. The sequence files are indicated by arrows with the names of sequence files. J_{LB} indicates the junction site between LSC region and IR_B region.

Each line indicates 10 kb long.

Table 2. List of identified genes and open reading frames, and their loci on the chloroplast genome.

Gene		From	To	Length (bp)	Amino acid residue	M.W.	Comments
(A)							
atpB	-	55846	54368	1479	492	53179.3	88.4% (Spi) 62.8% (Eco)
atpE	-	54362	53955	409	135	15054.3	63.0% (Spi) 22.2% (Eco)
trnM(CAU)	+	53801	53874	74	-		94.6% (Tob)
*trnV(UAC)	-	53652	53051	602	-		91.7% (Tob)
(Exon 1)		53652	53616	37	-		
(intron)		53615	53086	530	-		
(Exon 2)		53085	53051	35	-		
ndh3	-	52877	52515	363	120	14188.7	30.8% (mit)
psbG	-	52524	51793	732	243	27609.6	62.1% (Mz)
ORF169	-	51742	51233	510	169	20084.8	
(B)							
rbcL	+	56355	57782	1428	475	52790.0	90.5% (Spi)
trnR(CCG)	+	57877	57950	74	-		
ORF316	+	58065	59015	951	316	35826.3	
ORF36b	+	59193	59303	111	36	4017.8	
ORF184	+	59525	60079	555	184	21533.1	
ORF434	+	60151	61455	1305	434	51866.2	
petA	+	61641	62603	963	320	33482.6	78.8% (Spi)
(C)							
ORF40	-	62916	62794	123	40	4101.8	72.6% (Cya)
ORF38	-	63152	63036	117	38	4479.1	
psbF	-	63293	63174	120	39	4468.3	89.7% (Spi)
psbE	-	63554	63303	252	83	9493.5	89.2% (Spi)
ORF42a	-	63684	63556	129	42	5101.9	
(D)							
ORF31	+	64152	64247	96	31	3466.4	
ORF37	+	64370	64483	114	37	4075.9	
trnW(CCA)	-	64626	64553	74	-		93.4% (Spi)
trnP(UGG)	-	64788	64715	74	-		93.4% (Spi)
ORF42b	+	65027	65155	129	42	4746.6	
rpl33	+	65273	65470	198	65	7782.1	36.9% (Eco)
rps18	+	65498	65725	228	75	8879.5	34.7% (Eco)
(E)							
rpl20	-	66157	65807	351	116	12773.0	46.6% (Eco)
*rps12A	-	67057	?	?	123	13797.0	70.2% (Eco) 91.9% (Tob)
(Exon 1)		67057	66944	111	38		trans-split
(intron ?)		66943	?	?	-	(Exon 2,3 coded opposit strand)	
*ORF203	-	68640	67130	1511	203	22685.0	
(Exon 1)		68640	68570	71	-	(87.0% homologous to spinach X-gene)	
(intron)		68569	68052	518	-		
(Exon 2)		68051	67760	292	-		
(intron)		67759	67379	381	-		
(Exon 3)		67378	67130	249	-		
(F)							
psbB	+	69026	70552	1527	508	56191.5	88.2% (Spi)
ORF35	+	70669	70776	108	35	3958.8	
ORF27	+	70872	70955	84	27	3268.0	
ORF74	+	71092	71316	225	74	7928.3	
*petB	+	71424	72566	1143	215	24306.5	86.0% (Spi)
(Exon 1)		71424	71429	6	-		
(intron)		71430	71924	495	-		
(Exon 2)		71925	72566	642	-		
*petD	+	72715	73690	976	160	17413.5	95.6% (Spi)
(Exon 1)		72715	72722	8	-		
(intron)		72723	73215	493	-		
(Exon 2)		73216	73690	475	-		

Table 2 (continued).

Gene		From	To	Length (bp)	Amino acid residue	M.W.	Comments
(G)							
rpoA	-	74824	73802	1023	340	39240.2	25.6% (Eco), 54.1% (Spi)
rps11	-	75249	74857	393	130	14172.5	51.5% (Eco), 72.3% (Spi)
secX	-	75413	75300	114	37	4521.5	62.2% (Eco), 86.5% (Spi)
infA	-	75686	75450	237	78	8978.4	56.4% (Eco), 60.3% (Spi)
rps8	-	76171	75773	399	132	14921.4	45.5% (Eco)
rpl14	-	76621	76253	369	122	13496.6	58.2% (Eco)
*rpl16	-	77685	76719	967	143	16149.8	53.8% (Eco), 72.0% (Spir)
(Exon 1)		77685	77677	9	-		
(intron)		77676	77142	535	-		
(Exon 2)		77141	76719	423	-		
rps3	-	78396	77743	654	217	25055.0	40.6% (Eco)
rpl22	-	78804	78445	360	119	13580.8	37.8% (Eco)
rps19	-	79100	78822	279	92	10553.3	63.0% (Eco), 83.7% (Spi)
*rpl2	-	80514	79137	1378	277	31162.9	48.4% (Eco), 58.5% (Spi)
(Exon 1)		80514	80118	397	-		
(intron)		80117	79574	544	-		
(Exon 2)		79573	79137	437	-		
rpl23	-	80825	80550	276	91	10768.5	29.7% (Eco)
trnI(C*AU)	-	81057	80984	74	-		93.2% (Spi) C*; modified base
trnV(GAC)	+	81814	81885	72	-		95.8% (Spi) in IR region

Amino acid sequence homologies are calculated as $\frac{\text{(identical residue number)}}{\text{(residue number of liverwort product)}}$.

Homology percentages with gene products of Spinach (Spi), E. coli (Eco), Tobacco (Tob), human mitochondria (mit), Maize (Mz), Cyanella (Cya), and Spirodela oligorhiza (Spir) are shown in comments.

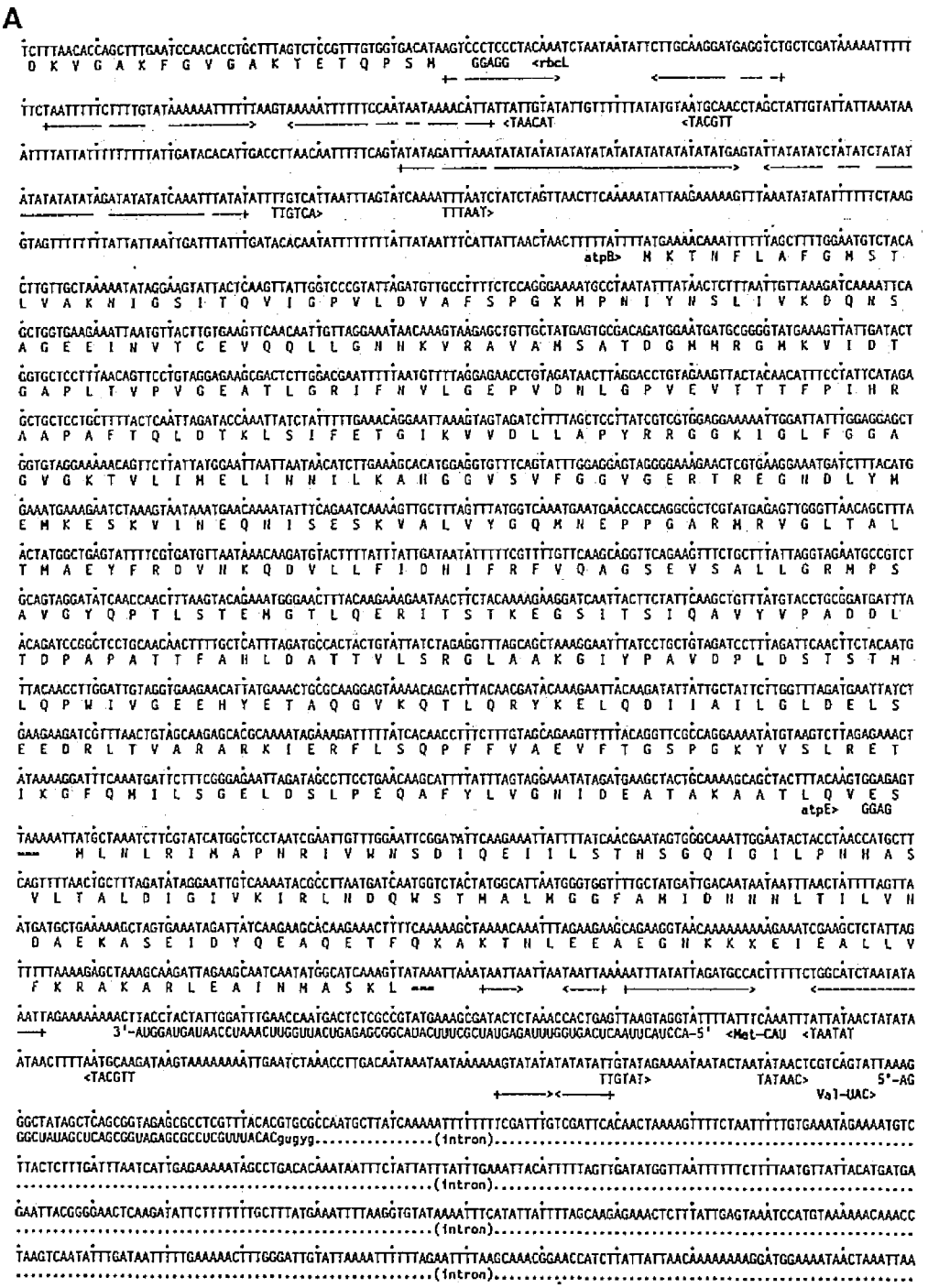

A
GGAGG
<rbcL
<TAACAT
<TACGTT
TTGTCA>
TTTAAT>
atpB>
atpE>
GGAG
<TACGTT
TTGTAT>
TATAAC>
5'-AG
Val-UAC>

Figure 4A.

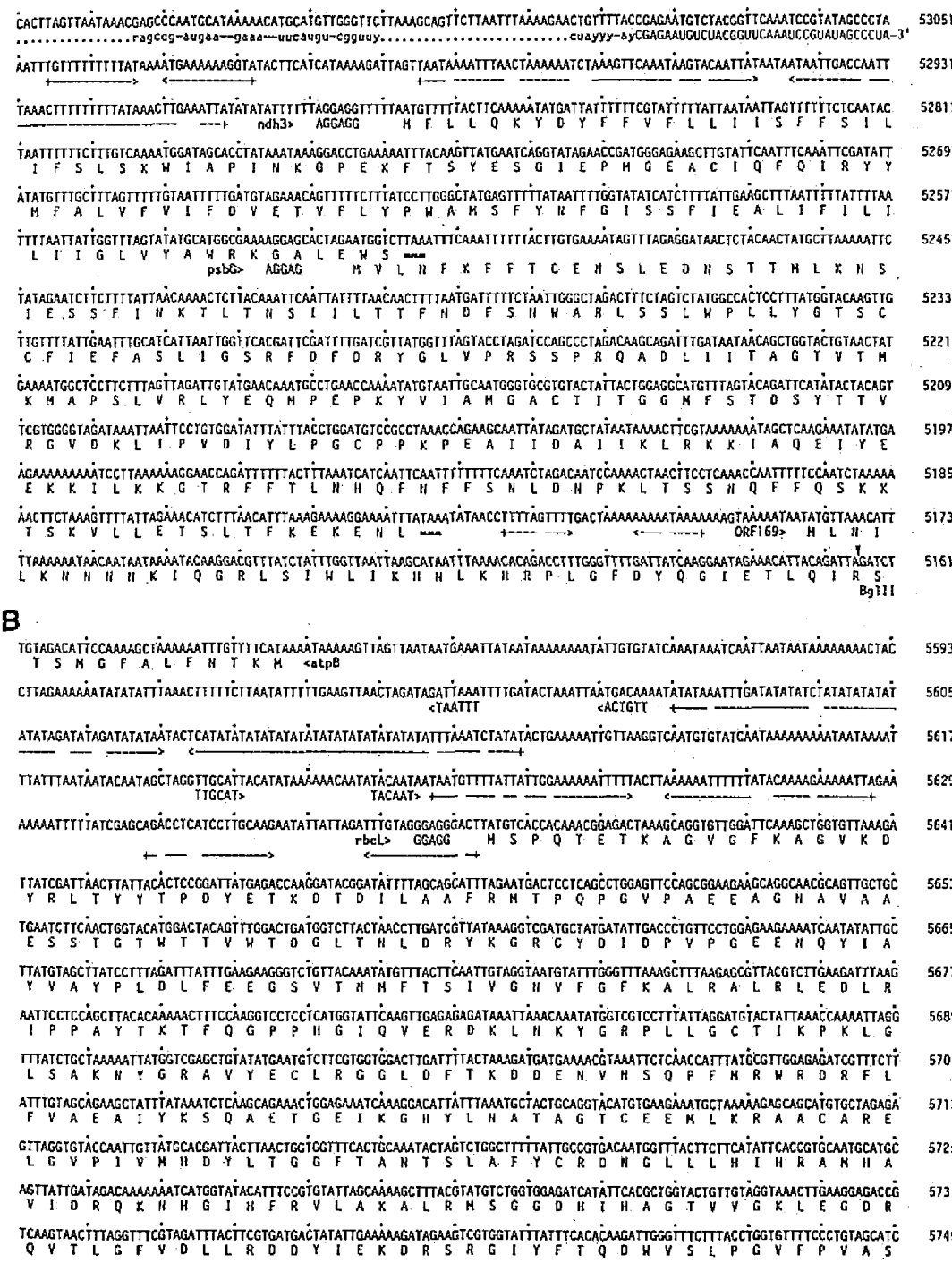


Figure 4A and 4B.

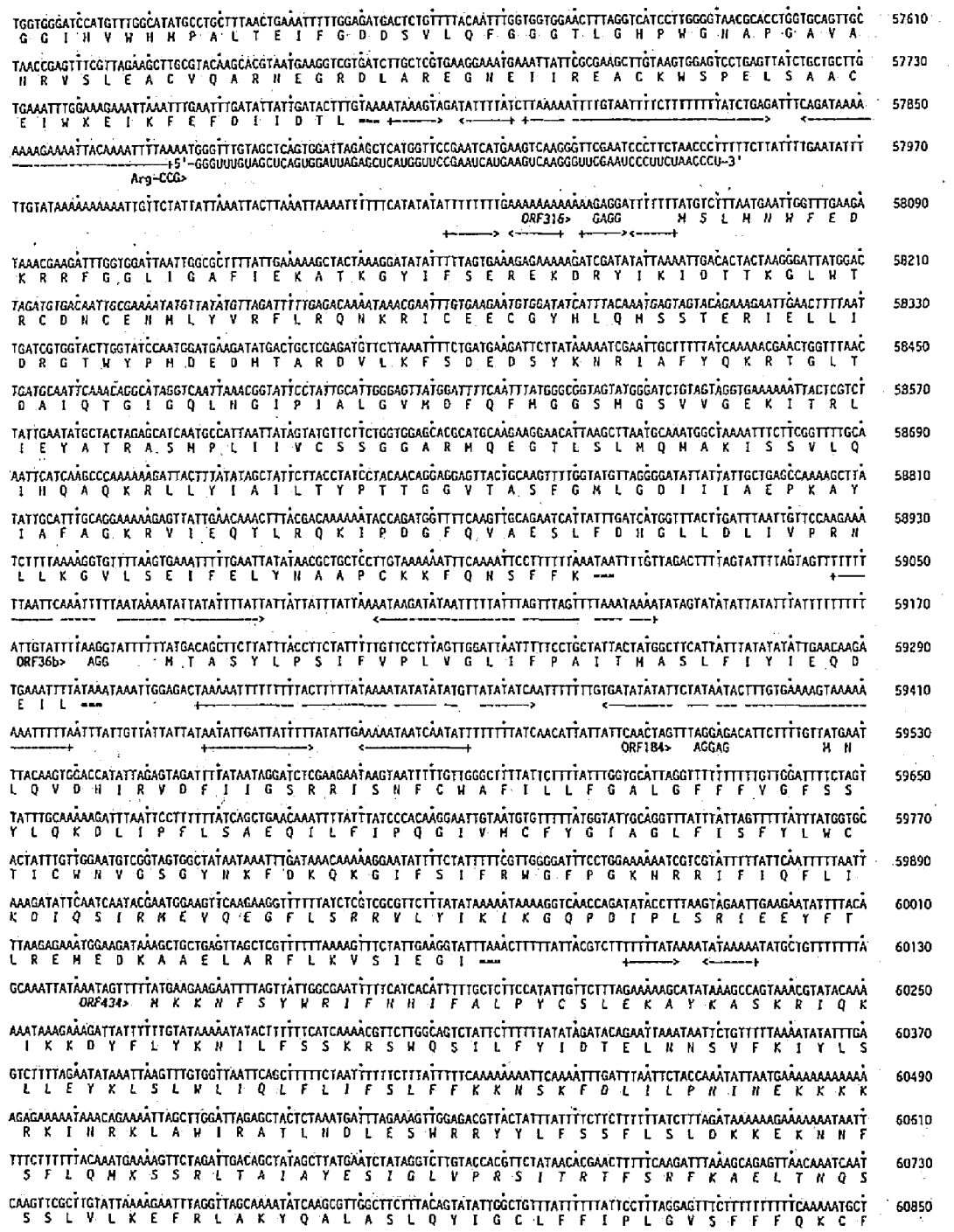


Figure 4B (continued).

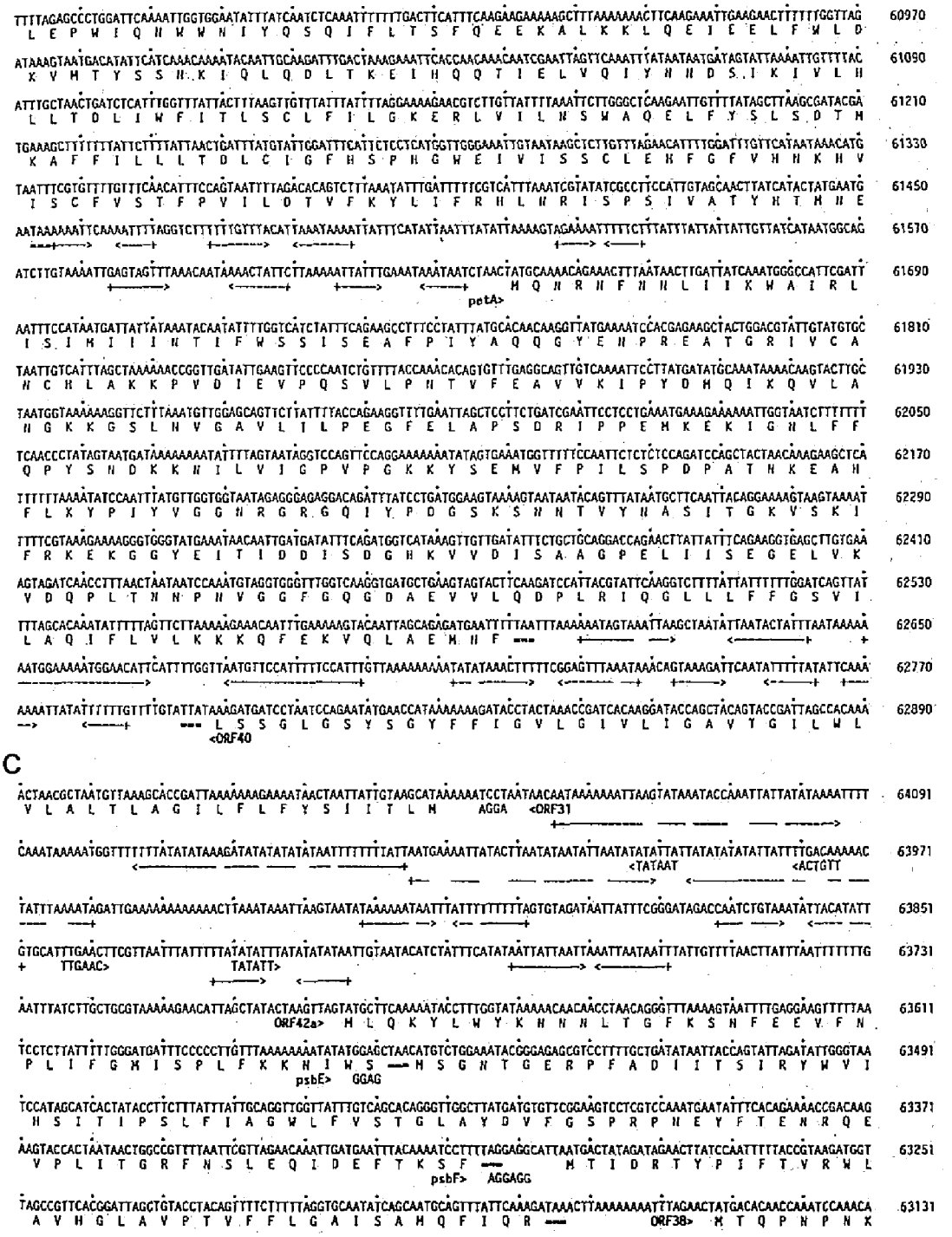


Figure 4B and 4C.

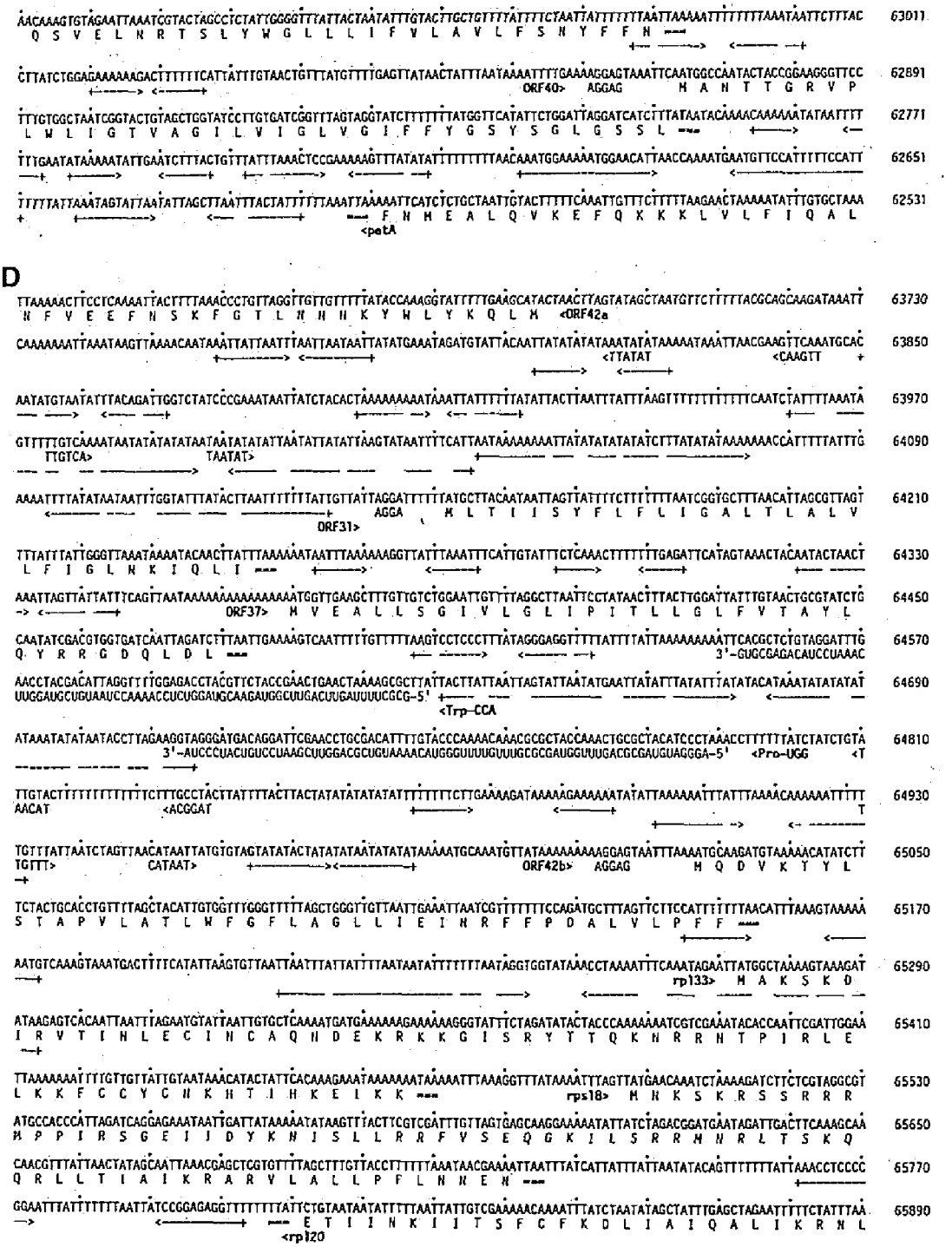


Figure 4C and 4D.

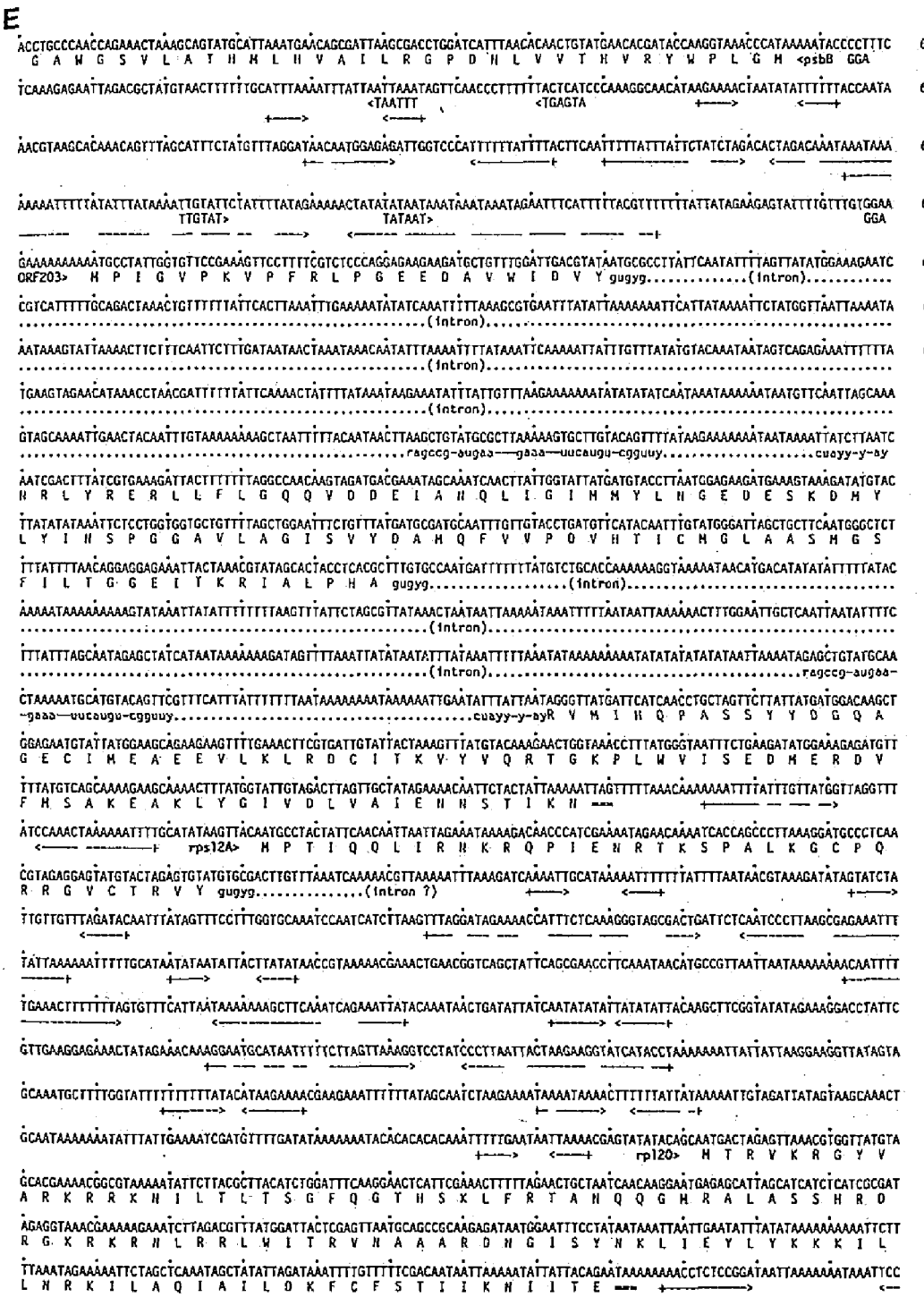


Figure 4E.

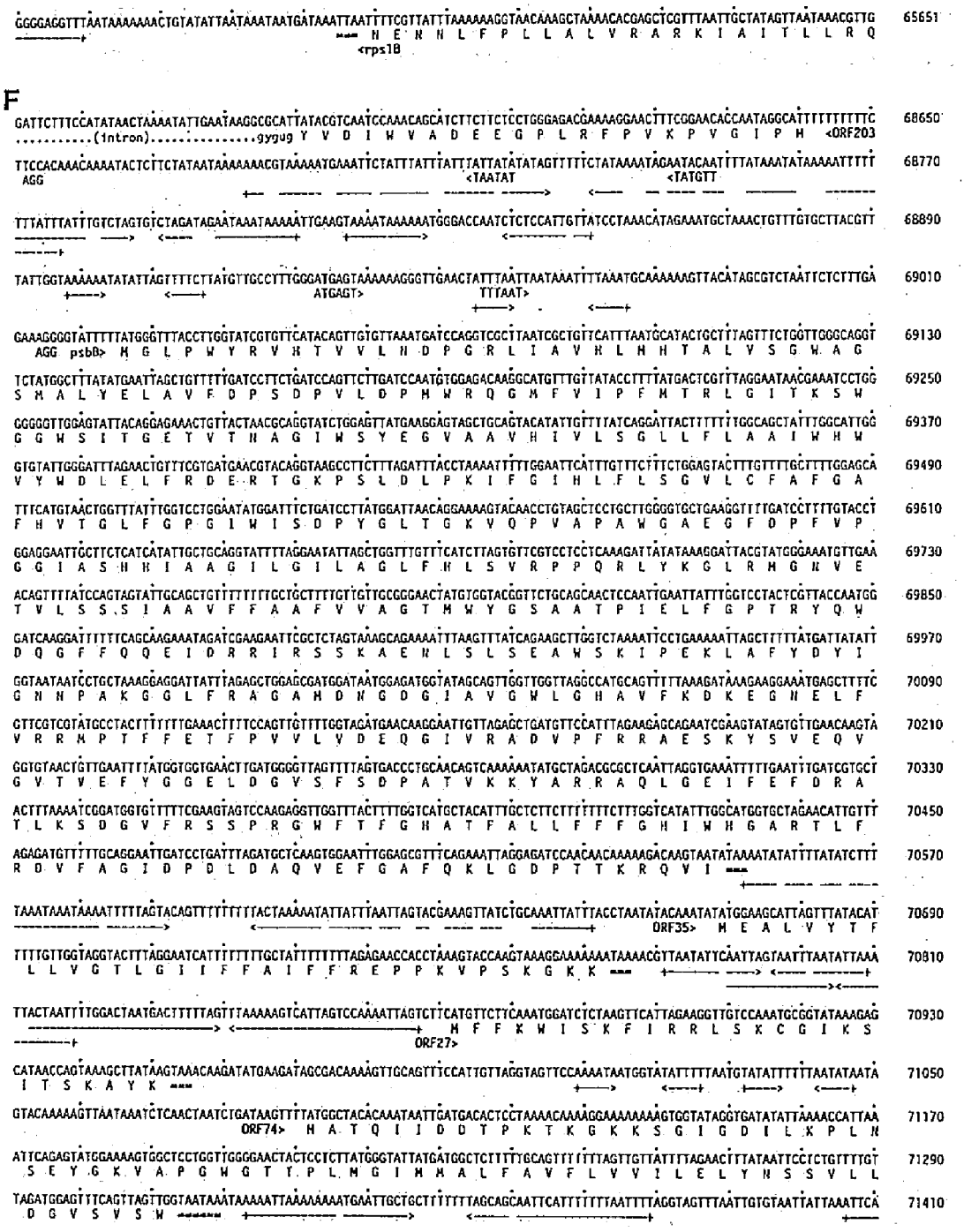


Figure 4E and 4F.

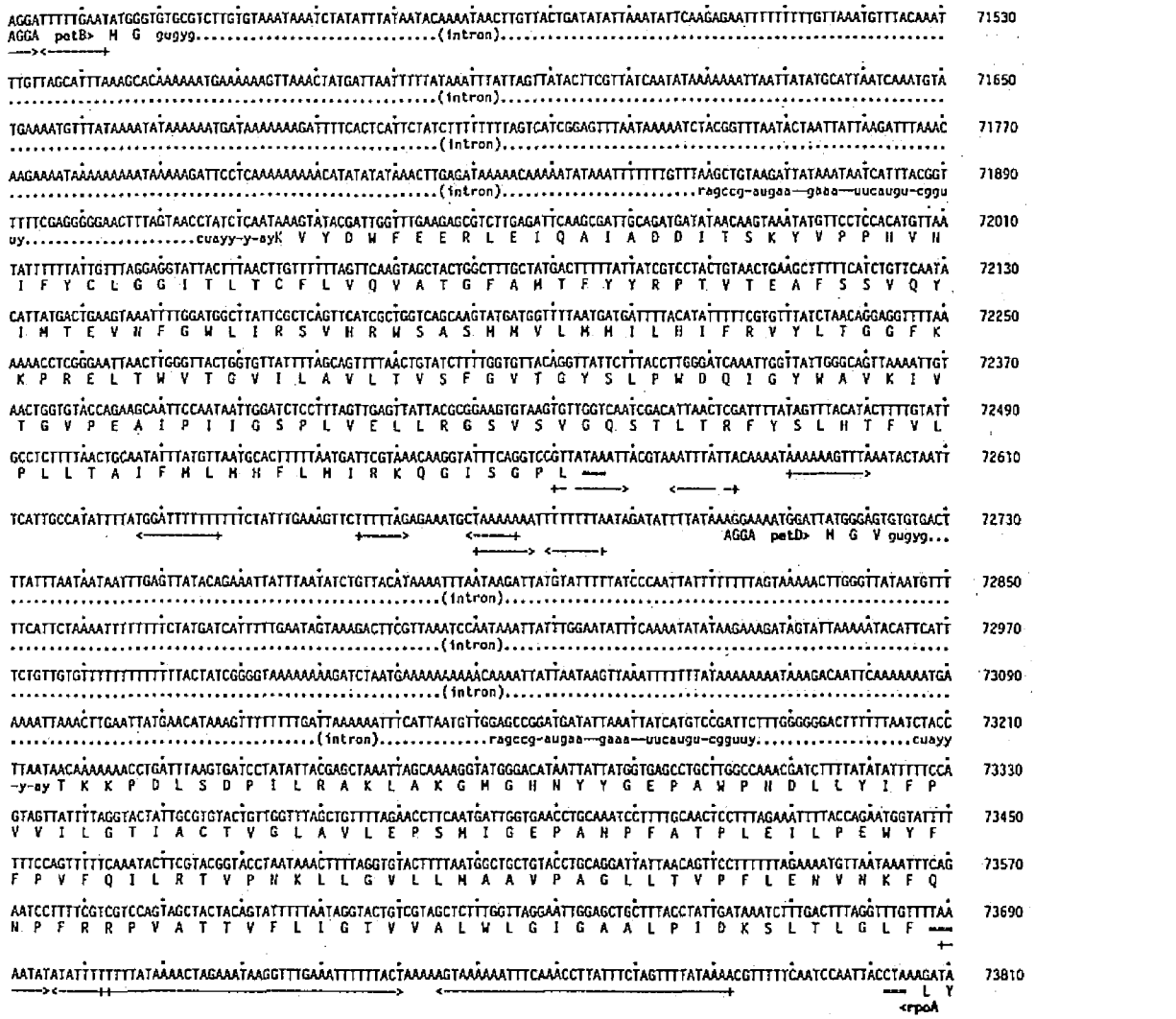


Figure 4F and 4G.

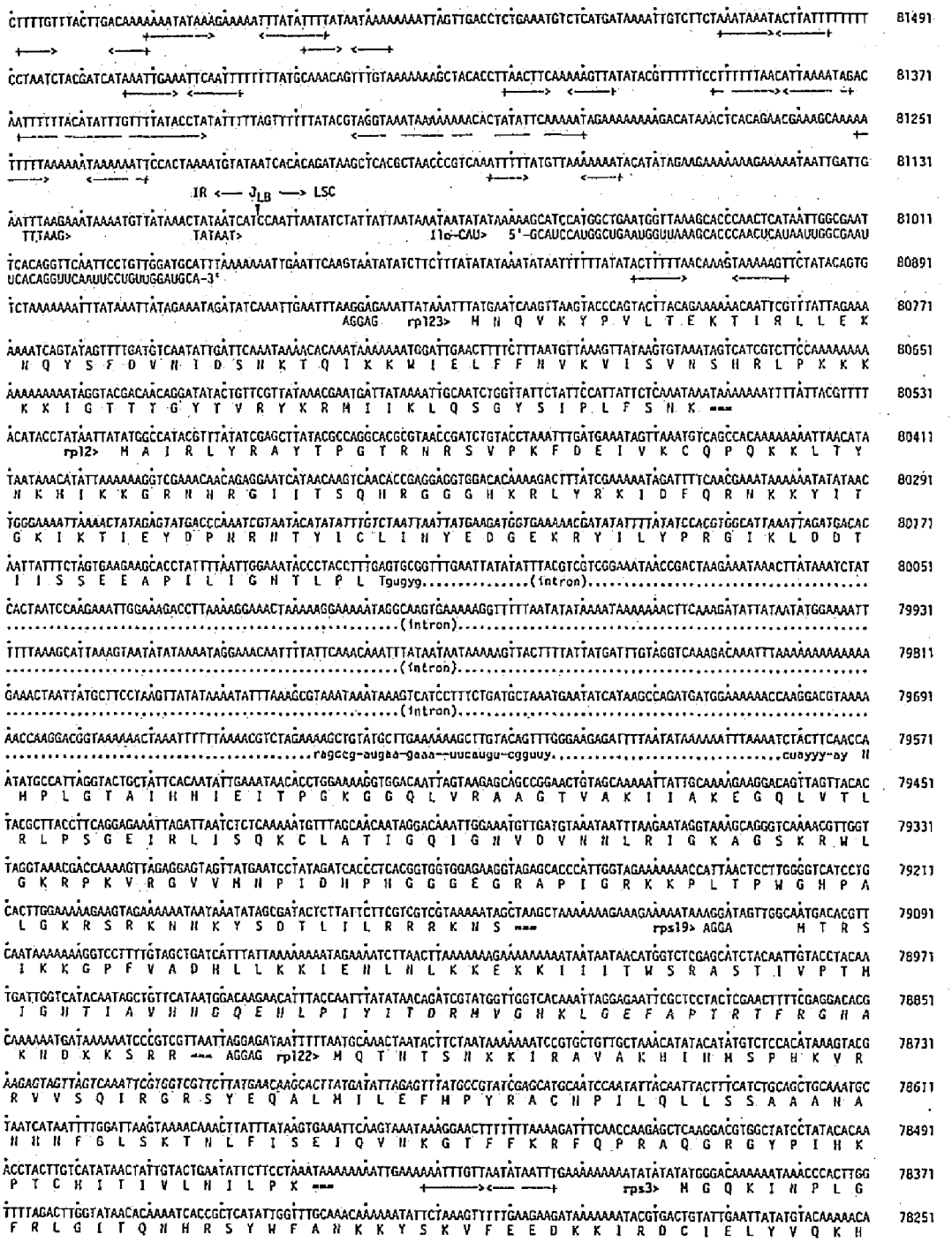


Figure 4G (continued).

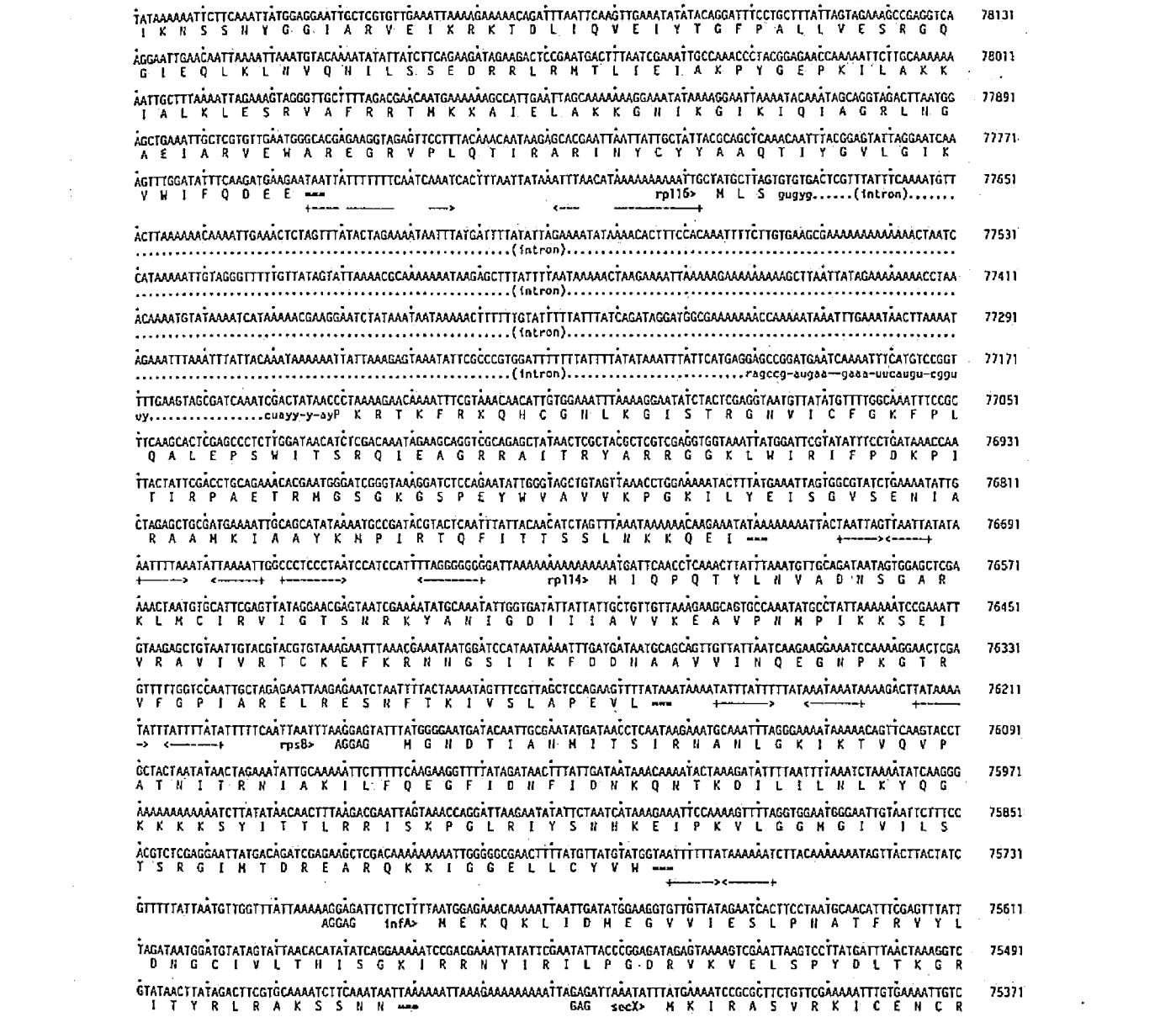


Figure 4G (continued).

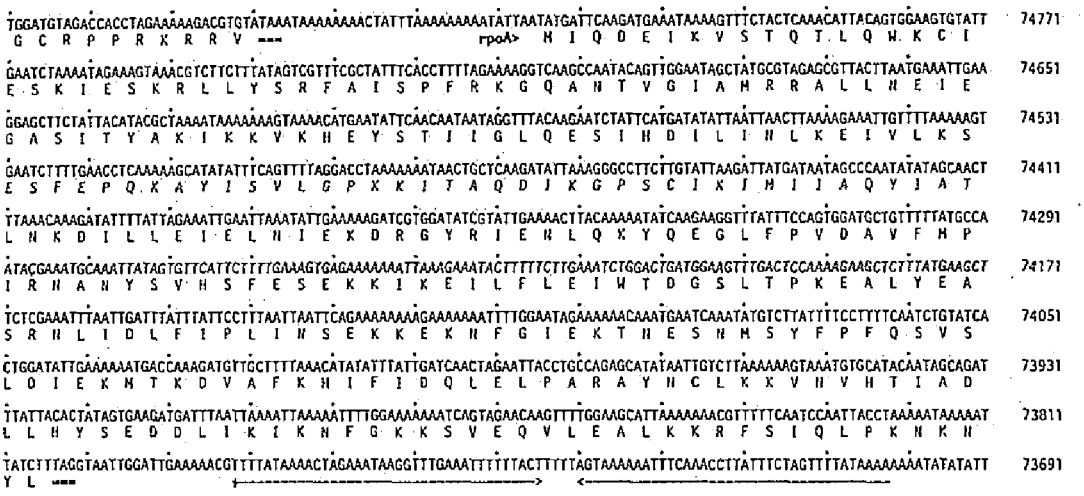


Figure 4G (continued).

Figure 4. Nucleotide sequence of the chloroplast DNA. The nucleotide position numbers are counted from the 5'-terminal nucleotide next to the inverted repeat IR_A. Dots are put on every ten nucleotide. Amino acid sequences deduced from the nucleotide sequences are shown under the DNA sequence by one letter symols. Stop codons are shown by double underlines. Putative stem-loop structures are shown by broken lines with arrow heads. Predicted promoter sequences and Shine-Dalgarno sequences are shown under the DNA sequence. Transfer RNA sequences are shown under the DNA sequence. Introns are shown by dots under the DNA sequence with 5' and 3' terminal consensus sequences (gagyg and ragccg.augaa..gaaa..uucaugu.cgguuy; r represents a or g, and y represents c or u).

nucleotide sequences of each block are shown in Fig. 4 A-G. The amino acid sequences of ORFs and the nucleotide sequences of transfer RNAs deduced from the DNA sequence are shown below the nucleotide sequences. Introns in ORFs are predicted in one tRNA gene (valine-UAC) and six protein coding sequences for petB, petD, rpl2, rpl16, and rps12 genes and ORF203 by the presence of the 5' consensus sequence (GUGYG; Y represents C or T), and 3' consensus secondary structures with the common sequences (RAGCCG.AUGAA..GAAA..UUCAUGU.CGGUUY; R represents A or G) characteristic to group II introns found in fungal mitochondrial genes and Euglena gracilis chloroplast genes (Michel and Dujon 1983, Keller and Michel 1985). Identified genes and open reading frames (ORFs), and their loci on the chloroplast genome are summarized in Table 2. Genes are categorized into three groups: II-1 transfer RNA genes; II-2 genes for photosynthetic polypeptides; II-3 genes for ribosomal proteins and subunits of RNA polymerase. In the section II-4, unidentified open reading frames are discussed. Detail characterization of these genes are described following sections.

II-1 Transfer RNA genes

As previously mentioned, chloroplasts contain genes for their own rRNAs. They probably also contain genes for all of their tRNAs. They show high homology with the corresponding genes of E. coli. Genes for numerous tRNAs have been sequenced and mapped on chloroplast chromosomes (Crouse et al. 1985).

RESULTS

From the DNA sequence, tRNA genes were predicted as regions that have higher GC content than spacer regions between ORFs as shown in Fig. 5. Seven tRNA genes were located by searching for the T-ψ loop consensus sequence (GTTCRA) and identified by constructing the clover-leaf structures as shown in Fig. 6.

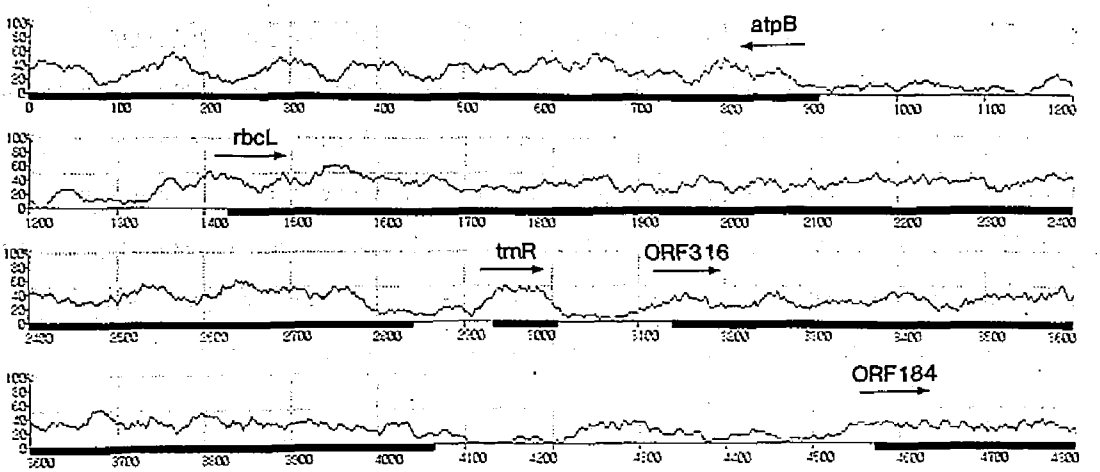


Figure 5. GC content and corresponding genes. GC content was plotted by calculation in average 30 nucleotides. The coding sequences are shown by bold lines with names of genes.

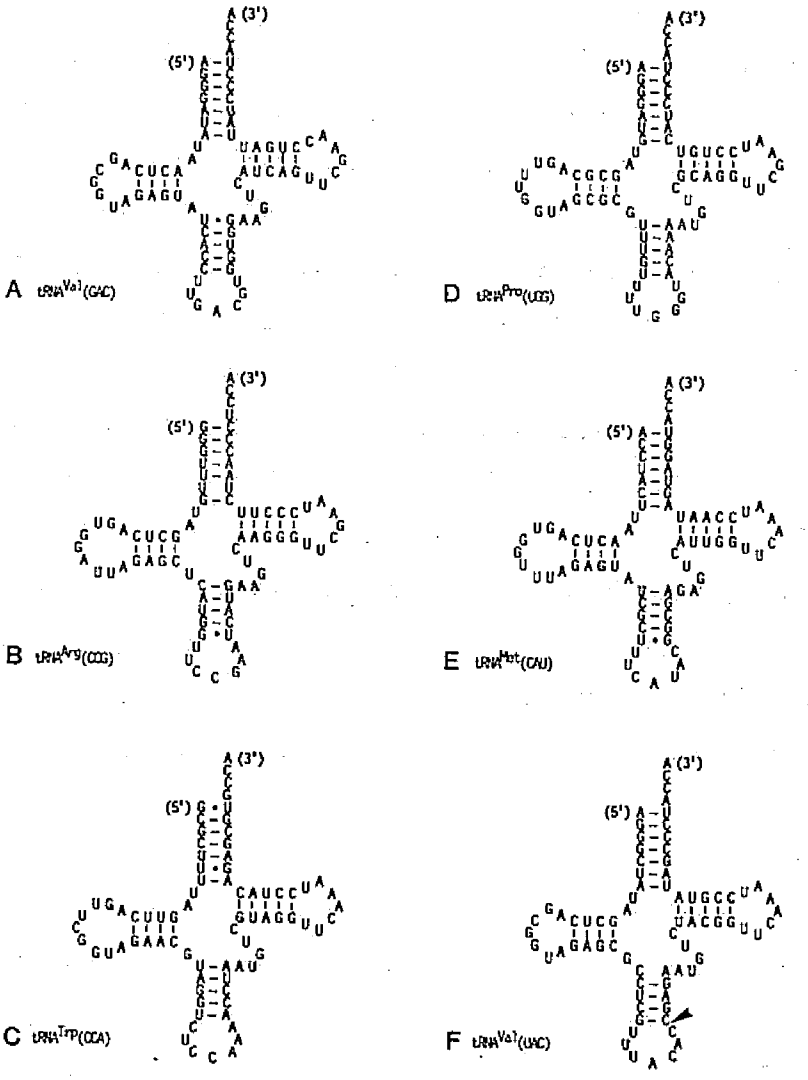


Figure 6. Secondary structures of tRNAs deduced from the DNA sequences.
The 3'-terminal CCA nucleotides are not coded by the chloroplast genome.
The insertion site of an intron in the $tRNA^{Val}$(UAC) is shown by an arrow head.

Valine and isoleucine tRNA genes (trnV-GAC and trnI-C*AU)

A gene for $tRNA^{Val}$(GAC) was found at position 81814 to 81885. The 5' terminus of 16S ribosomal RNA gene was mapped at the position 82109 by comparing with maize and tobacco chloroplast 16S rRNA genes (Schwarz and Kossel 1980, Tohdoh and Sugiura 1982). It was shown that the primary transcript of the 16S rRNA does not include $tRNA^{Val}$(GAC) (Strittmatter et al. 1985). In the liverwort putative promoter sequences were individually located upstream from 16S rRNA gene and valine tRNA gene. A tRNA gene (81057-80984) is located on the opposite DNA strand 756 bp apart from the 5' terminal end of the trnV-GAC gene. The unmodified anticodon is CAU complementary to the methionine codon AUG. The nucleotide sequence shows 43.2% and 54.1% homologies with the spinach initiator methionine tRNA(CAU) (Calagan et al. 1980) and elongator methionine tRNA(CAU) (Pirtle et al. 1982), respectively, but exhibits 93.2% homology with the spinach chloroplast isoleucine tRNA(C*AU). In addition, the tRNA gene in liverwort has extra mismatching within the anticodon stem as seen in the case of spinach isoleucine tRNA(C*AU) (Kashdan MA et al. 1982 and Francis et al. 1982). Therefore, coding sequence for this tRNA can be tRNA gene (trnI-C*AU) highly modified in the first nucleotide of the anticodon.

Arginine tRNA gene (trnR-CCG)

A tRNA gene (57877-57950) was found 94 bp downstream from the termination codon of the rbcL gene. The tRNA has the anticodon of CCG that can recognize CGG arginine codon. A pair of mismatching nucleotides (U-U) was found in the amino acyl stem (see Fig. 6B). The liverwort trnR-ACG gene in the IR region also have two mismatching nucleotides (U-U and U-U) in its amino acyl stem (Kohchi et al. 1986).

Tryptophan and proline tRNA genes (trnW-CCA and trnP-UGG)

Coding sequences for two tRNAs were found at positions 64626-64553 and 64788-64715 near the psbE gene in Fig. 4D. Their secondary structures showed the anticodon triplets, CCA and UGG, pairing with codons for tryptophan UGG and proline

CCA, respectively (Fig. 6C and 6D). Therefore these tRNAs were identified to be tryptophan tRNA and proline tRNA. The $tRNA^{Trp}$(CCA) and $tRNA^{Pro}$(UGG) show 93.4% and 93.4% sequence homologies with spinach chloroplast $tRNA^{Trp}$(CCA) and $tRNA^{Pro}$(UGG) (Canaday et al.1981 and Francis et al. 1982), respectively. The genes for these tRNAs were designated trnW-CCA and trnP-UGG. Their coding sequences were separated by 88 bp spacer region. A significant promoter sequence for these genes is present 20 bp upstream from the 5'end of the proline tRNA gene, but not in the spacer region between two tRNA genes. Two tRNA genes must be co-transcribed in a primary transcript and processed into mature tRNA molecules, although 42 bp stem structure in the spacer region can be formed as described in $tRNA^{Arg}$(CCG).

Elongator methionine and valine tRNA genes (trnM-CAU and trnV-UAC)

In the BglII fragment Bg10, 83 bp apart from atpE coding region, a tRNA gene (53801-53874) was found on the opposite strand. The anticodon of the tRNA was AUG. The tRNA showed sequence homology with spinach elongator methionine tRNA (94.8%), and with initiator methionine tRNA (46.8%). Therefore this tRNA gene was confirmed to be elongator methionine tRNA gene. A tRNA gene (53652-53051), whose anticodon was UAC was found 148 bp apart from trnM-CAU. This putative valine tRNA gene was split by 530 bp intron at the junction between anticodon stem and loop as shown in Fig. 6F.

DISCUSSION

Thirty two species of tRNA genes have been identified and mapped on the liverwort chloroplast genome (Ohyama et al. 1986). Those identified tRNAs are listed in the codon table as shown in Table 3. The tRNAs encoded by chloroplast genome are sufficient to read all codons taking into account an extended wobbling and modification in the anticodons. No tRNA gene would encode a 3'-terminal CCA nucleotides. Five species of tRNA genes (trnI-GAU, trnV-UAC, trnA-UGC, trnK-UUU and trnG-UCC) have been found to be split by group II introns. A tRNA gene (trnL-UAA)

Table 3. Codon table and unmodified anticodons of tRNAs coded by Marchantia polymorpha chloroplast genome.

codon		anticodon	codon		anticodon	codon		anticodon	codon		anticodon
UUU	Phe		UCU	Ser		UAU	Tyr		UGU	Cys	
UUC		GAA	UCC		GGA	UAC		GUA	UGC		GCA
UUA	Leu	UAA*	UCA		UGA	UAA	Ter		UGA	Ter	
UUG		CAA	UCG			UAG	Ter		UGG	Trp	CCA
CUU	Leu		CCU	Pro		CAU	His		CGU	Arg	ACG
CUC			CCC		(GGG)	CAC		GUG	CGC		
CUA		UAG	CCA		UGG	CAA	Gln	UUG	CGA		
CUG			CCG			CAG			CGG		CCG
AUU	Ile	GAU*	ACU	Thr		AAU	Asn		CGU	Ser	
AUC			ACC		GGU	AAC		GUU	AGC		GCU
AUA		CAU	ACA		UGU	AAA	Lys	UUU*	AGA	Arg	UCU
AUG	Met	CAU	ACG			AAG			AGG		
	fMet	CAU									
GUU	Val		GCU	Ala		GAU	Asp		GGU	Gly	
GUC		GAC	GCC			GAC		GUC	GGC		GCC
GUA		UAC*	GCA		UGC*	GAA	Glu	UUC	GGA		UCC*
GUG			GCG			GAG			GGG		

AUG codon is an initiation codon. Termination codons (UAA, UAG and UGA) are indicated by Ter. Asterisks indicate presence of introns in the coding sequences.

is interrupted by group I intron. Five species of tRNA genes (trnV-GAC, trnI-GAU, trnA-UGC, trnR-AGC, trnN-GUU) have been identified in each inverted repeat region (see Fig. 2).

Three kinds of arginine tRNA genes have been identified on the chloroplast genome. Duplicated genes, trnR-ACGs are localized in IR regions (Kohchi et al. 1986) and a trnR-UCU gene was mapped near the 3' side of the atpA gene coding for α subunit of H^+-ATP synthase (Umesono et al. 1986, see Fig. 2). In liverwort chloroplast genome, however, the trn^{Arg}(CCG) gene, which has not been found in the chloroplasts of any species of plants, was identified near the 3' end of the rbcL gene. Arginine codons in the codon table are separated into two boxes including AGR and CGN codons (see Table 3). Codons, AGA and AGG, can be read by $tRNA^{Arg}$(UCU) using G/U wobbling of the third letter. The CGU and CGG codons can also be recognized by $tRNA^{Arg}$(ACG) and $tRNA^{Arg}$(CCG), respectively. However, CGC and CGA codons could not be read by $tRNA^{Arg}$(ACG) and $tRNA^{Arg}$(CCG) without modification of the first letter of ACG and CCG anticodons. In E. coli there are also three species of arginine tRNAs having anticodons ACG, CCG and UCU. Therefore these results indicate that the mechanisms of the codon-anticodon recognition in chloroplasts is similar to those in E. coli. In addition, unpaired nucleotides in amino acyl stem were also reported in chloroplast $tRNA^{Arg}$(ACG) molecules of liverwort (Kohchi et al. 1986), Euglena gracilis (Hallick et al. 1984), Spirodela oligorhiza (Keus et al. 1984) Pelargonium zonale (Hellmund et al. 1984), and tobacco (Kato et al. 1985). These incomplete structures of amino acyl stems may alter the codon-anticodon recognition. A promoter sequence is not detected in the 94 bp spacer region between rbcL and trnR coding regions. Instead, there are two stem structures consist of 8 and 32 bp long (dG= -44.9 kcal) that may function as intracistronic termination or RNA processing signals (Fig. 4B). If it is so, the trnR gene may be co-transcribed with rbcL gene and processed into mature tRNA.

II-2 Genes for photosynthetic polypeptides

The most important role of chloroplasts is photosynthesis. Liverwort chloroplast genome is smaller (121 kb) than those in other higher plants (Palmer et al. 1985). Nevertheless it is interesting to know whether the liverwort chloroplast genome has necessary sets of genes for photosynthetic polypeptides. Events in photosynthesis take place on the thylakoid membrane as well as in stroma of the chloroplasts. The thylakoid membrane contains reaction centers (called photosystems I and II), cytochrome b6/f complex and coupling factors. A cytochrome b6/f complex in chloroplasts operates in an electron transfer chain of a plastoquinol-plastocyanin oxidoreductase between photosystem II and I. The complex has four major polypeptides. Three of them, cytochrome f, cytochrome b_6 and 17 kd subunit 4, are synthesized in the chloroplasts, whereas the Rieske FeS protein is synthesized in the cytoplasms (Hauska 1985). The genes for cytochrome b_6 and the 17 kd subunit 4 are separated by 1 kb on the chloroplast genome and are transcribed as a common precursor mRNA (Alt et al. 1983). The gene for cytochrome f is located distantly from others and translated into a preprotein larger than mature cytochrome f, suggesting that processing occurs during insertion from the ribosomes in the stroma into the thylakoid membrane (Alt et al. 1983, Whilley et al. 1984, Alt and Herrmann 1984).

RESULTS

Genes for photosynthetic polypeptides were identified by comparing amino acid sequences of the ORFs with those of photosynthetic proteins reported previously in other species of plants. The amino acid sequence alignments of the identified photosynthetic proteins are shown in Fig. 7 A-K. On the LSC region sequenced in this study, there are ten genes for photosynthetic proteins; the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (rbcL), 51 kd P-680 chlorophyll a apoprotein (psbB), cytochrome b-559 polypeptides (psbE and psbF), cytochrome f preprotein (petA), cytochrome b6 (petB), cytochrome b6/f complex subunit 4 (petD)

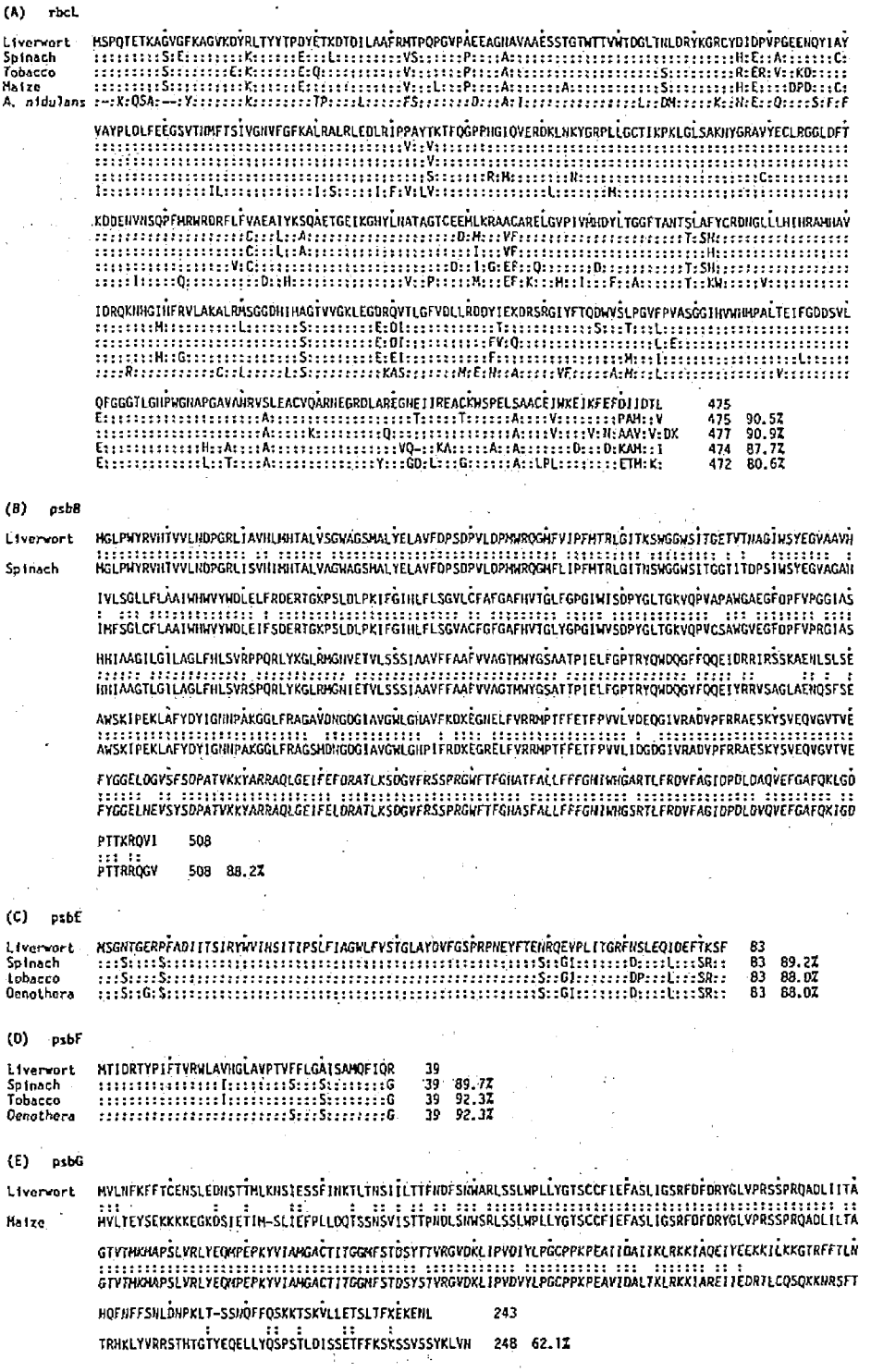


Figure 7 (continued).

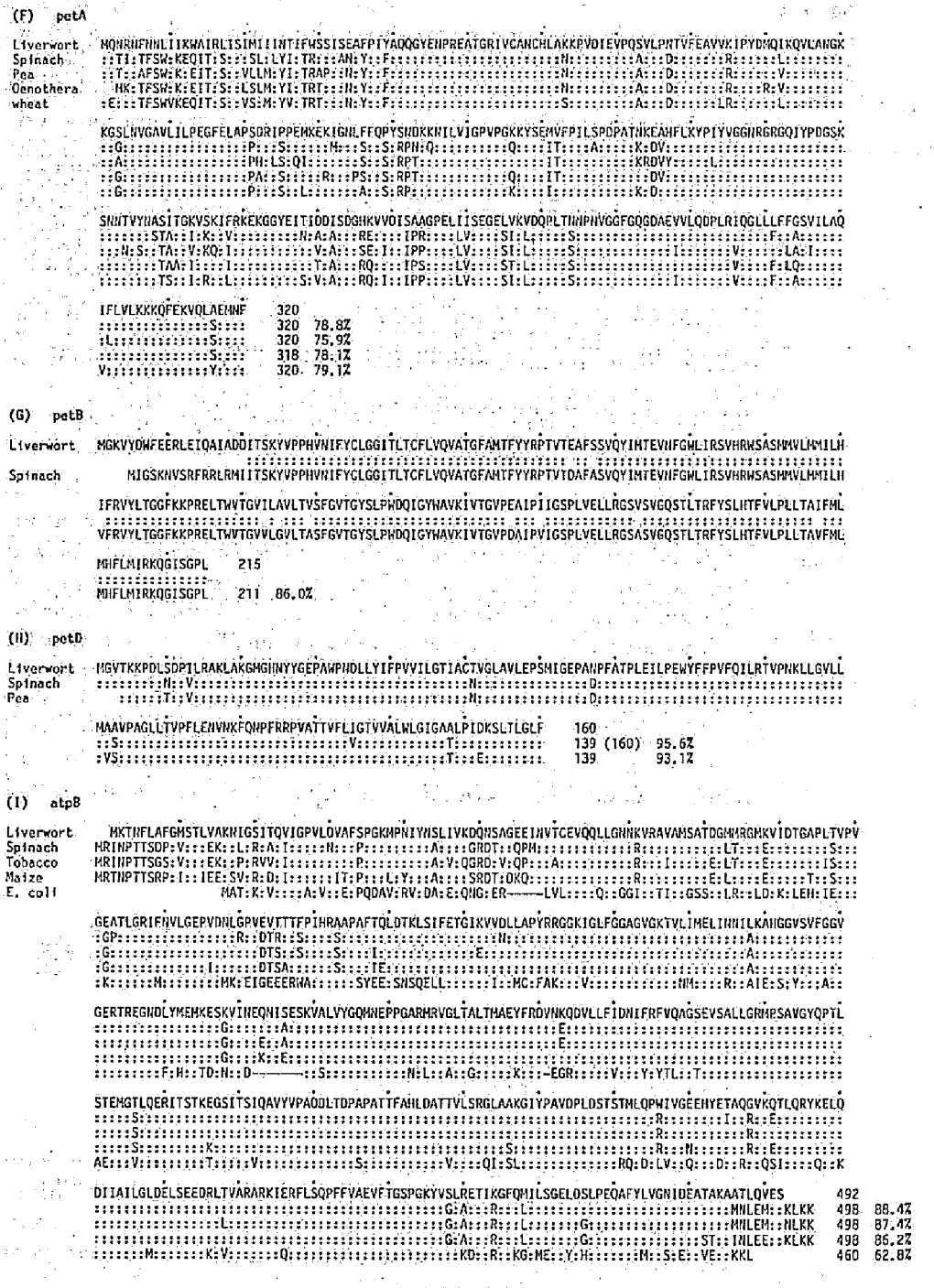


Figure 7 (continued).

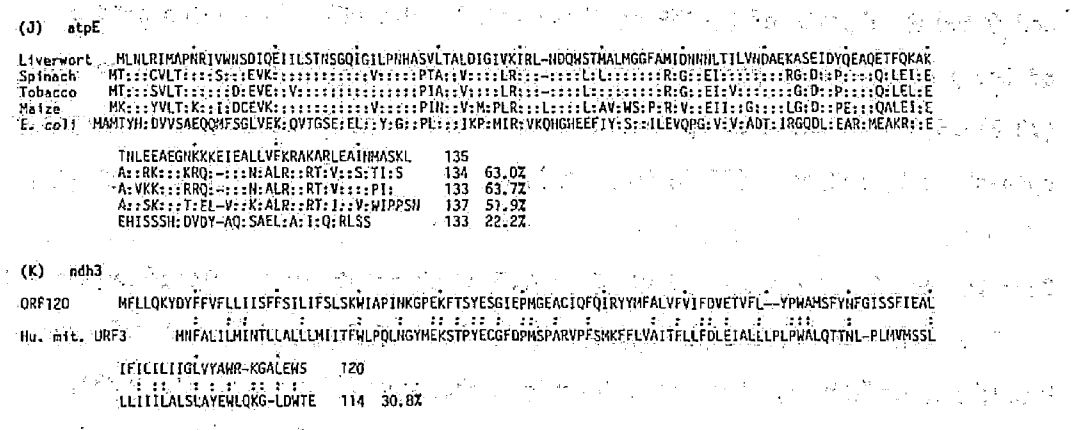


Figure 7. Amino acid sequence alignments of photosynthetic polypeptides. The amino acid sequences are shown by one letter codes. Identical amino acid residues are shown by colons, and deleted residues are shown by dashes. The amino acid residue numbers and sequence homologies with liverwort gene products are indicated at the end of sequences.

and β and ϵ subunits of H^+-ATP synthase (atpB and atpE). The amino acid sequences of the photosynthetic proteins in liverwort chloroplasts exhibit high homologies (78.8%-95.6%) to those of spinach chloroplasts except for atpE (63.0% homologous to spinach atpE) and psbG (62.1% to maize).

Gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (rbcL)

The rbcL gene was previously mapped in the Bam5 fragment by heterologous hybridization with the tobacco chloroplast rbcL gene (Ohyama et al. 1983). An ORF475 (56355-57782) shows high amino acid sequence homology with rbcL of spinach (90.5%) (Zurawski et al. 1981), tobacco (90.5%) (Shinozaki and Sugiura 1982), maize (87.6%) (McIntosch et al. 1980) and A. nidulance (80.6%) (Shinozaki and Sugiura 1983) (Fig. 7A). The rbcL gene was localized 508 bp apart on the opposite strand to the atpB gene coding for β subunit of H^+-ATP synthase. In the spacer region between atpB and rbcL genes, typical stem-loop structures can be formed including 45 bp AT-rich stem structure (dG= -44.9 kcal).

Genes for photosystem II P-680 chlorophyll a apoprotein (psbB), and for cytochrome b-559 polypeptides (psbE and psbF)

The amino acid sequence of an ORF508 (69026-70552) shows 88.2% homology to that of the spinach 51 kd photosystem II P-680 chlorophyll a apoprotein (Morris and Herrmann 1984). The amino acid sequence alignment with that of spinach is shown in Fig. 7B.

ORF83 (63554-63303) and ORF39 (63293-63174) are located close together (10 bp apart). The ORF83 shows 89.2% amino acid sequence homology to that of the spinach apocytochrome b-559 polypeptide, which is psbE gene product (Herrmann et al. 1984). The ORF39 shows 89.7% amino acid sequence homology to the b-559 URF39 gene product of spinach (Herrmann et al. 1984). The psbE and psbF genes are proceeded by the typical Shine-Dalgarno-like sequences (GGAGG and AGGAGG, respectively). The psbF gene terminates by a TAG stop codon that overlap to two nucleotides of the SD

sequence for psbE. Their sequence alignments are shown in Fig. 7C and 7D.

Gene for photosystem II G-protein (psbG)

Recently Steinmetz et al. identified a new protein (248 amino-acid residues long) associated with the photosystem II complex and analyzed the fine structure of its gene, psbG, on the maize chloroplast genome (1986). The M. polymorpha counterpart of psbG gene (52524-51793; encoding 243 amino acids, 27.6 kd) was identified and the predicted amino acid sequences were compared as shown in Fig. 7E. Unlike other photosystem II polypeptides, the psbG proteins are significantly diverse in N- and C-terminal portions; in average they are only 62.1% homologous, whereas the central portions (maize, amino acid residue number 36-182; M. polymorpha 37-183) are 91.9% identical. In M. polymorpha, the psbG gene overlaps with the last seven nucleotides of the preceding ndh3 gene.

Gene for cytochrome f preprotein (petA)

An ORF320 (61641-62603) shows high degree of amino acid sequence homologies to those of the cytochrome f preprotein of spinach (78.8%) (Alt et al. 1984), pea (75.9%) (Willey et al. 1984), Oenothera hookeri (78.1%) (Tyagi et al. 1986), and wheat (79.1%) (Willey et al. 1984). The ORF320 can be liverwort cytochrome f preprotein gene (petA) whose molecular weight is 33.4 kd. The amino acid sequence alignments compared with those of other plant species are shown in Fig. 7F. The N-terminal 35 amino acid sequence of liverwort cytochrome f polypeptides shows relatively lower homology (40.0%) to that of spinach. On the other hand, the remaining sequence of 285 residues gives high homology (83.5% with spinach cytochrome f mature protein). However, the hydrophobic characteristics of the N-terminal region are conserved in petA gene products indicating that N-terminal 35 amino acid residues may be functional as a signal peptide. The molecular weight of liverwort mature cytochrome f polypeptide is 31.3 kd.

Genes for cytochrome b6 (petB) and for cytochrome b6/f complex subunit 4 (petD)

ORF162 (72078-72566) can be postulated by using initiation codon AUG (72078-72080). The C-terminal amino acid sequence was found to be homologous to that of the spinach cytochrome b6 polypeptide (petB) (Heinemeyer et al. 1984). However the N-terminal portion including initiation codon is quite different from that of spinach petB gene. Instead, a consensus sequence for the 3' end of the intron (RAGCCG.AUGAA..GAAA..UUCAUGU.CGGUUY) was found in the beginning (71859-71924) of the ORF162. A consensus sequence specific to the 5' end of the intron (GUGYG) was also located further upstream of the ORF162. This indicates that the liverwort petB gene (71424-72566) is interrupted by 495 bp intron (71430-71924) and codes a polypeptide of 215 amino acid residues (Fig. 4F and 7G).

The amino acid sequence of ORF139 (73271-73690) shows 95.6% and 93.1% homologies to cytochrome b6/f complex subunit 4 of spinach (Heinemeyer et al. 1984) and pea (Phillips et al. 1984), respectively (Fig. 7H). However, the consensus sequences for an intron (493 bp: 72723-73215) are also present in petD gene as in the petB gene. The putative secondary structures of introns in petB and petD genes are shown in Fig. 8. The initiation codon of the petD gene can be extended to the upstream methionine codon (72715) as shown in Fig. 4F. The molecular weight of the liverwort petD gene product can be estimated to be 17.4 kd (160 amino acid residues) considering presence of an intron in the coding sequence. On the other hand, intron specific sequences can also be identified on the DNA sequences of the spinach and pea genes, although they do not describe the presence of introns in the coding sequences for petD gene. If this is true, the amino acid sequences of the 5' extended region of the petD genes were highly conserved in the three species of plants (95.6% homologous to spinach petD protein, see Fig. 4F). A sequence motif (AGGA) for the ribosome binding sites are present 9 bp upstream from initiation codons of each petD and petB gene.

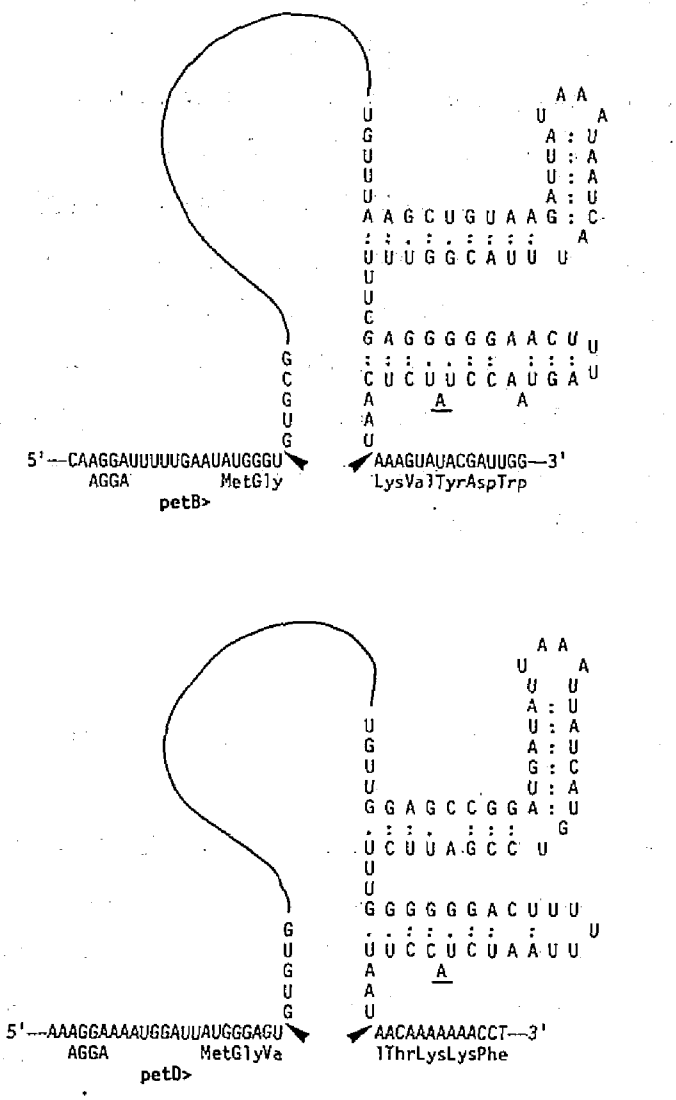


Figure 8. Secondary structures of introns in petB and petD precursor mRNA molecules. The Shine-Dalgarno sequences and amino acid sequences are shown under the messenger RNA molecules. Underlined nucleotide A is putative branch point in the lariat form. The junction sites between exons and introns are indicated by arrow heads.

Genes for β and ε subunits of H^+-ATP synthase (atpB and atpE)

A gene cluster, atpB (55846-54368) - atpE (54362-53955), was localized 508 bp apart from rbcL in head to head on the opposite DNA strand. The nucleotide-binding subunit β encoded for by atpB gene contains 492 amino acids (Higgins et al. 1985). The M. polymorpha atpB amino acid sequence showed 88.4% homology to spinach atpB gene product (Zurawski et al. 1982), 86.2% to maize (Krebbers et al. 1982), 62.8% to E. coli (Saraste et al. 1981). In contrast, M. polymorpha ε subunit (135 amino acid residues), product of atpE, is less homologous to that of spinach (63.0%, Zurawski et al. 1982), maize (51.9%, Krebbers et al. 1982) and E. coli (22.2%, Saraste et al. 1981). Although the Shine-Dalgarno sequence of atpE gene was overlapped with the atpB coding region in spinach and maize, the coding sequences of these two genes of M. polymorpha were not overlapped as shown in Fig. 9.

DISCUSSION

The comparison of amino acid sequences of photosynthetic polypeptides between liverwort and other higher plant chloroplasts reveals the highly conserved amino acid sequences (78.8-95.6% with spinach photosynthetic polypeptides) during the events of chloroplast evolution. In contrast, it is interesting to note less amino acid sequence conservation in transcriptional and translational polypeptides including α subunit of RNA polymerase (54%), ribosomal proteins (46-68%) and initiation factor 1 (65%). These results suggest that chloroplast ribosomal proteins and RNA polymerases may have changed faster than photosynthetic polypeptides in the event of evolution.

Previous studies showed that several plants have two rbcL mRNA transcripts (Erion 1985). It has been shown that the smaller mRNA transcripts are produced by processing of the primary rbcL transcripts (Mullet et al. 1985). Although the spacer region between rbcL and atpB coding regions show less sequence homology with those of other plant species, several regions are conserved in land plant

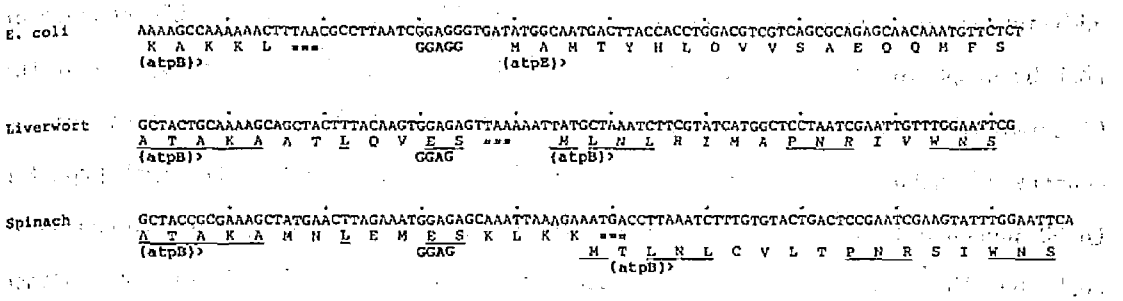


Figure 9. Comparison of nucleotide sequence of spacer region between atpB and atpE genes. The identical amino acid residues between M. polymorpha atpBE and spinach atpBE proteins are underlined. Shine-Dalgarno sequences are shown under the nucleotide sequences by bold letters.

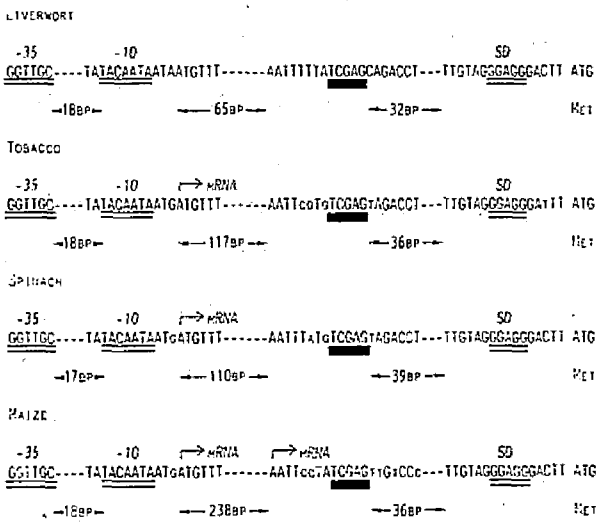


Figure 10. Comparison of the leader sequences of rbcL gene. The promoter sequences referred to "-10" and "-35", and Shine-Dalgarno sequences are double underlined. The conserved sequences in the 5' leader regions of plant chloroplast rbcL gene are indicated by thick underlines.

chloroplasts. The promoter sequences for rbcL and atpB referred to "-10" and "-35" regions, and the Shine-Dalgarno sequence (GGAGG) for rbcL were conserved. In addition, a sequence motif (TCGAG) was found 38-45 bp apart from the SD sequence of rbcL by comparing the DNA sequences as shown in Fig. 10. Interestingly, one of the 5' terminal sites of processed rbcL mRNAs are located 8-12 bp upstream from this conserved sequence motif in maize and spinach. The distance from the rbcL promoter to SD sequence in liverwort rbcL is shorter than those in other plant chloroplast rbcL, but the distances between promoter regions and this newly identified sequence motifs were not conserved in land plants. This sequence motif may have some roles on the translational regulation of rbcL gene expression.

In the liverwort cytochrome f preprotein (petA), the N-terminal region (1-35 amino acid residue) shows rather low homology (40%) with corresponding regions of those from pea and spinach chloroplasts. The N-terminal 35 amino acid residues of the pea and spinach cytochrome f proteins have been shown to be removed as a signal peptide for localization into the thylakoid membrane (Willey et al. 1983).

The liverwort chloroplast petB and petD genes were found to be interrupted by introns classified as group II intron in fungal mitochondria (Keller and Michel 1985). Introns were not reported in the spinach chloroplast petB and petD genes and it has been suggested that the psbB, petB and petD genes are co-transcribed in a primary transcript following intercistronic splicing (Westhoff 1985). A similar gene organization as in the liverwort genome can be interpreted in the spinach nucleotide sequence presented by Heinemeyer et al. (1984) by application of intron consensus sequences (Keller and Michel 1985, Ohyama et al. 1986). The molecular weight calculated from the deduced amino acid sequences of the petD (17.4 kd) agrees reasonably with that (17.5 kd) of the spinach subunit 4 polypeptide determined by SDS-polyacrylamide gel electrophoresis (Alt et al. 1983). Findings of introns in liverwort petB and petD genes and absence of promoter sequences and termination signals in the spacer regions between psbB, petB and petD genes suggest that these three genes are transcribed in a 7.3 kb common precursor mRNA. Further study to

identify the splicing junction between their exons is in progress.

II-3 Genes for ribosomal proteins and α subunit of RNA polymerase (rpoA)

Gene expression in the chloroplasts shows many similarities with those in prokaryotic cells such as E. coli. Ribosomal RNAs (rRNA), transfer RNAs (tRNA) and ribosomal proteins involving the transcription and translation in chloroplasts have prokaryotic features (Whitfeld and Bottomley 1983, Kozak 1983). Chloroplast ribosomes play a fundamental role not only in chloroplast biogenesis but also in photosynthesis, since several photosynthetic proteins are made on chloroplast ribosomes (Ellis 1981). Chloroplast ribosomes share many structural and functional similarities in common with prokaryotic ribosomes including subunit sedimentation coefficients, mode of action, size and sequences of rRNAs (Bartsch et al. 1982, Shmidt et al. 1985, Bartsch 1985). The similarities in the translational systems are so extensive that initiation and elongation factors, even ribosomal subunits of E. coli and chloroplast can be interchanged (Graves and Spremulli 1983, Schmidt et al. 1983). Some of the cDNA coding for the chloroplast ribosomal protein have been obtained from pea (Gantt and Key 1986).

The genes for chloroplast ribosomal proteins; S4 (Subramanian et al. 1983), S7 (Montandon and Stutz 1984), S11 (Muller et al. 1986), S12 (Montandon and Stutz 1984), S16 (Shinozaki et al. 1986), S19 (Sugita and Sugiura 1983), L2 (Zurawski et al. 1984), S14 (Umesono et al. 1984, Kirsch et al. 1986), and L16 (Posno et al. 1986) have been identified by amino acid sequence homology with the corresponding E. coli ribosomal proteins. One third of chloroplast ribosomal proteins are thought to be encoded by chloroplast genome (Filho et al. 1981) but some of the rests are shown to be encoded by nuclear genome as a precursor polypeptides (Schmidt et al. 1985). It is not so far clarified how many genes for ribosomal proteins are encoded by chloroplast genome.

Two different types of RNA polymerase activities appear to be present in chloroplasts (Greenberg et al. 1984). One is associated with a transcriptionally

active chromosome and is preferentially active in rRNA synthesis (Briat and Mache 1980). The other is readily extracted in soluble form active in tRNA and mRNA transcription (Gruissem et al. 1983). The site of synthesis of these RNA polymerases have not proved to be either in cytoplasms or in chloroplasts (Lerbs et al. 1985, Muller et al. 1986). The soluble RNA polymerase from maize (Kidd and Bogorad 1979) and pea (Tewari and Goel 1983) has been shown to have subunit structure.

RESULTS

Thirteen ORFs shows significant amino-acid sequence homologies with respective E. coli ribosomal proteins (L2, L14, L16, L20, L22, L23, L33, S3, S8, S11, S12, S18, and S19). The amino acid sequence alignments with corresponding E. coli ribosomal proteins are shown in Fig. 11 A-M. Percentages of the amino acid sequence homology of liverwort ORFs to E. coli ribosomal protein genes range from 25.3% (L23) to 70.2% (S12).

A large cluster of ribosomal protein genes was found on the LSC region next to the Junction site (J_{LB}) (Fig. 3G and 4G). The organization of the gene cluster was (trnI)-L23-35bp-L2-36bp-S19-17bp-L22-48bp-S3-57bp-L16-97bp-L14-81bp-S8-86bp-(infA)-36bp-(secX)-50bp-S11-32bp-(rpoA). The putative infA gene product exhibits 56.4% and 60.3% amino acid sequence homologies to the E. coli initiation factor 1 (Pon et al. 1979) and the spinach infA product (Muller et al. 1986), respectively (Fig. 11-O). The secX is the putative gene for unknown polypeptide expressed in E. coli (Ceretti et al. 1983). The amino acid sequence shows 62.2% homology to the E. coli secX protein and also shows 86.5% homology to the spinach open reading frame corresponding to the liverwort secX protein (Muller et al. 1986) (Fig. 11P). The rpoA is the putative gene for α subunit of RNA polymerase whose amino acid sequence shows homologies to those of E. coli (25.6%, Meek et al. 1984) and spinach (54.1%, Muller et al. 1986) (Fig. 11N). Spacer regions between their genes in the cluster appears to be less than 97 bp long. No significant prokaryotic promoter sequences

Figure 11 (continued).

Figure 11 (continued).

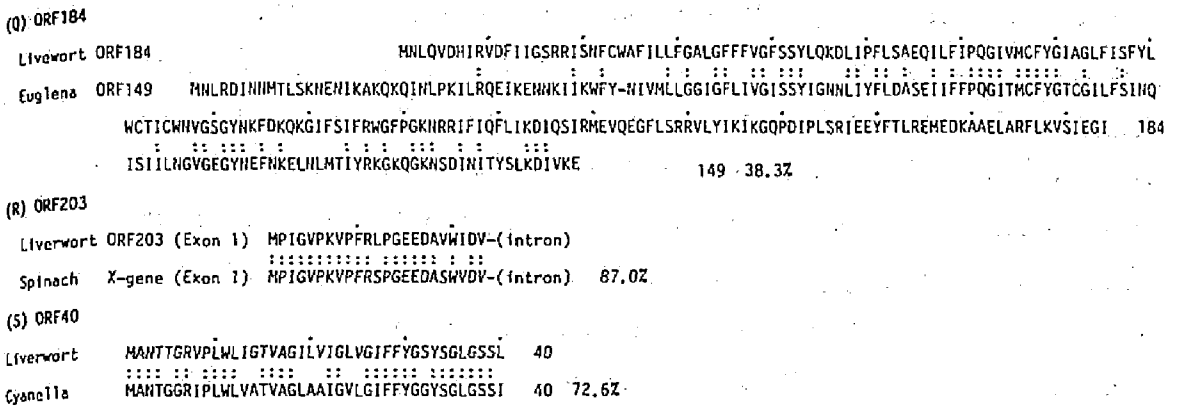


Figure 11. Amino acid sequence alignments of ribosomal related proteins and subunit of RNA polymerase. The amino acid sequences are shown by one letter codes. Identical amino acid residues are shown by colons, and deleted residues are shown by dashes. The amino acid residue numbers and sequence homologies with liverwort gene products are indicated at the end of sequences.

were observed between them. Another ribosomal protein gene cluster is localized between trnP-UGG and psbB as the order of trnP-124bp-L33-27bp-S18-81bp-L20-786bp-S12(exon1)-72bp-ORF203-385bp-psbB. The rps12 gene (Exon1) is localized 72 bp downstream from TAG stop codon of ORF203. The rest of the exons are located far apart on the different DNA strand indicating trans-splicing mechanism for the gene expression (Fukuzawa et al. 1986). Detail discussion on the trans-split gene rps12 are in Chapter III. The coding sequences for ribosomal protein L2 and L16 are interrupted by 545 bp and 534 bp group II introns, respectively (Fig. 4F).

DISCUSSION

In the liverwort chloroplast genome, nucleotide sequence revealed a large cluster of genes coding for ribosomal and related proteins (trnI-L23-L2-S19-L22-S3-L16-L14-S8-infA-secX-S11-rpoA) on the LSC region near the J_{LB}. Upstream from the gene for L23 ribosomal protein, there is an isoleucine tRNA(C*AU) gene whose promoter highly functioned in E. coli as well as in chloroplasts (Fukuzawa et al. 1985). The length of the gene cluster from the 5' end of isoleucine tRNA gene to 3' end of rpoA gene is approximately 7.3 kb. No promoter sequence can not be found in the short (less than 97 bp) spacers between ribosomal protein genes indicating that the gene cluster may be transcribed into a single precursor RNA from the trnI to rpoA genes. Furthermore, this cluster has similar order to the clusters reported in the E. coli ribosomal protein operons such as the S10 operon (Zurawski and Zurawski 1985), spc operon (Cerretti et al. 1983), and alpha operon (Bedwell et al. 1985) (S10-L3-L4-L23-L2-S19-L22-S3-L16-L29-S17, L14-L24-L5-S14-S8-L6-L18-S5-L30-secY-secX, and S13-S11-S4-rpoA-L17, respectively). Two additional clusters of ribosomal protein genes are seen in the orders of S12(Exon1)-L20 and S18-L33 between trnP-UGG and psbB genes. The genes for S12 (from the 39th amino acid residue to C-terminal end) and S7 ribosomal protein also have been clustered as seen in E. coli str-operon (S12-S7-EF_G-EF_{Tu}, Post et al. 1978, Post and Nomura 1980). In contrast, the chloroplast rps2, rps4, rps14, rps15 and rpl21 genes are scattered throughout the

liverwort chloroplast genome.

The rpl2 genes of spinach and Nicotiana debneyi has been located in the inverted repeat regions. In N. debneyi, but not in spinach, rpl2 is interrupted by a 666 bp intron (Zurawski et al. 1984). However, liverwort rpl2 gene is located on the LSC region just outside of the IR_B region and has 544 bp intron. Introns are also present in the rpl16 gene of Spirodela oligorhiza (interrupted by 1411 bp intron) (Posno et al. 1986). The first exon of genes for L16 ribosomal protein encodes for the first three amino acid residues (Met-Leu-Ser) of N-terminal. It is interesting that the products of ribosomal protein genes encoded in the chloroplast genome are important components in the initial stage of ribosome and rRNA assembly (Dorne et al. 1984).

A putative gene for α subunit of RNA polymerase is located at the 3' end of the gene cluster for ribosomal proteins. Genes for β and β' subunits of RNA polymerase (rpoB, rpoC1 and rpoC2) are located as a single operon at 50 kb apart from rpoA gene (Umesono et al. 1986). An ORF homologous to the E. coli sigma subunit of RNA polymerase is not found on the liverwort chloroplast genome.

II-4 The putative gene ndh3 and unidentified open reading frames

It is reported that the M. polymorpha chloroplast genome contained a set of homologues of mammalian mitochondrial "URF" genes (Ohyama et al. 1986). In this region, a gene named ndh3 corresponding to human mitochondrial URF3 was identified by amino acid sequence comparison (Fig. 7K). The mitochondrial URF3 gene code for the component of respiratory-chain NADH dehydrogenase complex (Chomyn et al. 1985). An ORF120 (52877-52515) is located at the upstream region of psbG with an overlap of seven nucleotides. This ORF dose not contain an intron and the product (120 amino acid residues, 14.2 kd) is similar in size to a human mitochondrial URF3 protein (114 amino acid residues, Anderson et al. 1981) sharing 30.8% homology. Actually proteins of chloroplast ndh genes are not yet identified, but a 3' half of ndh3 gene would be also conserved in maize chloroplast genome; the published sequence of maize

psbG gene contains its 5' flanking region of 158 nucleotides (Steinmetz et al. 1986). If two G residues (located at -68 and -96 in their numbering) are deleted, the region from -157 to +7 will encode a polypeptide similar to the last 54 amino acids of the M. polymorpha ndh3 product (85.2% identical) as well as an overlapping to the downstream psbG gene. There has been no previous report of the presence of ndh genes in chloroplast genome. However, an NADH-plastoquinone-(PQ) oxidoreductase activity has been detected in the chloroplasts of Chlamydomonas reinhardii (Bennoun 1982), thus it is possible that this ORF encodes one of subunits of the NADH-PQ oxidoreductase.

Fifteen significant ORFs, which do not show any homologies with previously reported genes, were located on the sequenced region in this study (see Table 2).

Amino Acid sequences of two unidentified open reading frames in liverwort chloroplast genome show significant homologies with those of unidentified frames reported in other kinds of chloroplast genome. An ORF184 (59525-60079) shows 38.3% local homology to the ORF149 located at the next to the gene for elongation factor Tu of Euglena chloroplasts (Fig. 11Q). The Euglena ORF149 does not terminate of its stop codon but follow intron sequence (Montandon and Stutz 1983). It is interesting to compare the C-terminal region of liverwort ORF184 (position 128 to 184 amino acid residue) with the corresponding Euglena ORF in the Exon 4 described by them. The first exon (68640-68570) of the ORF203 shows 20 out of 23 amino acid identity with an reading frame in spinach X-gene on the opposite strand of psbB gene as shown in Fig. 11R. It is reasonable to believe that the open reading frames conserved in two kinds of chloroplasts would code polypeptide having an unknown function. In addition, an ORF40 (62916-62794) showed 72.6% homology with cyanella Cyanophora paradoxa ORF40, which is not proved to have any function (Fig. 11S, Bryant personal communication).

CHAPTER III The split gene for chloroplast ribosomal protein S12

Introns (intervening sequences) in a chloroplast RNA gene have been reported; the 23S rRNA gene of Chlamydomonas reinhardii (Rochaix and Malone 1978, Rochaix et al. 1985); the tRNA genes, trnI-GAU and trnA-UGC, in the 16S-23S rDNA spacer region of Zea mays (Koch et al. 1981) and Nicotiana tabacum (Takaiwa and Sugiura 1982); as well as the chloroplast tRNA genes trnL-UAA (Steinmetz et al. 1983a, Bonnard et al. 1984), trnK-UUU (Sugita et al. 1985), trnG-UCC (Deno et al. 1984a, Quigley and Weil 1985) and trnV-UAC (Deno et al. 1982, Krebbers et al. 1984, Zurawski and Clegg 1984). Introns within chloroplast protein genes also have been reported in several genes of Euglena gracilis; for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (rbcL) (Koller et al. 1984), the elongation factor Tu (tufA) (Montandon and Stutz 1983), and the 32-kd protein (psbA) (Karabin et al. 1984, Keller and Michel 1985). The gene for the 32-kd protein of C. reinhardii also has introns (Erickson et al. 1984) as does the gene for the H^+-ATP synthase subunit I (atpF) of wheat (Bird et al. 1985). Zurawski et al. (1984) reported that the chloroplast ribosomal protein L2 (rpl2) in Nicotiana debneyi has a single intron. Several genes in the chloroplast DNA from the liverwort, M. polymorpha are shown to have introns in their coding sequences (see Fig. 2 in chapter II).

Hallick et al. (1985) and Fromm et al. (1986) reported that the reading frame of the ribosomal protein S12 in N. tabacum is interrupted by two introns, but they described only the second one. During nucleotide sequencings of the chloroplast DNA from the liverwort, M. polymorpha, however, the first exon with the 5' intron boundary sequence was found on the opposite strand of the chloroplast DNA. In this chapter, the complex structure of the putative gene for chloroplast ribosomal protein S12 from the liverwort, M. polymorpha is presented, which has three exons split into different DNA strands. The mechanisms of the expression of this unusually organized gene will be discussed.

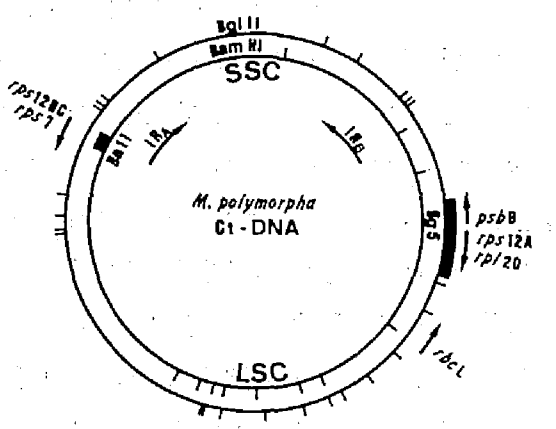


Figure 1. Locations of coding regions for chloroplast ribosomal protein S12 on physical maps of chloroplast DNA from a liverwort, M. polymorpha. Exon 1 (rps12A) was located on the BglII fragment (Bg5); and the direction of its transcription was clockwise indicated by an arrow. By contrast, exons 2 and 3 (rps12B and C) were found on the BamHI fragment (Ba11), and their transcription being in the opposite direction from that of exon 1 (counter clockwise). The abbreviations rps12, rps7 and rpl20 are for genes of ribosomal proteins S12, S7 and L20. The site of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (rbcL) is shown. IR_A and IR_B indicate a set of inverted repeats, and SSC and LSC indicate the small single copy and large single copy regions, respectively.

MATERIALS AND METHODS

Chloroplast DNA fragments, the BamHI (Ba11) and BglII (Bg5) fragments, were cloned into the respective plasmids, pBR322 and pKC7, as described by Maniatis et al. (1982). A physical map of the chloroplast DNA for BamHI fragments has been described previously (Ohyama et al. 1983, Umesono et al. 1984). The location of the Bg5 fragment on that map was determined by restriction analysis and Southern hybridization (Fig. 1). Methods for the sequence determination are described in chapter II.

RESULTS AND DISCUSSION

At first a coding region for ribosomal protein S12 on the BamHI fragment (Ba11) was identified using Southern hybridization with an E. gracilis probe (provided by Drs. Montandon and Stutz). The Ba11 fragment was mapped at the junction (J_{LA}) between the inverted repeat (IR_A) and the large single copy (LSC) region (Fig. 1). DNA sequence analysis of the Ba11 fragment revealed that the coding sequence for the ribosomal protein S12 was found, however, the N-terminal 38 amino acids of the protein was missing in this coding region.

By amino acid homology search between E. coli S12 (Post et al. 1978) protein and open reading frames deduced from the nucleotide sequence data files, the missing N-terminal 38 amino-acid sequence was found on the BglII fragment (Bg5) approximately 60 kb away on the opposite DNA strand (Fig. 1). Complete nucleotide sequences of the coding regions for ribosomal protein S12, rps12A and rps12B-C, including the flanking regions, are shown in Fig. 2. Exons 1 and 2 were followed by a consensus sequence (GTGCG) of the 5' boundary regions found in fungal mitochondrial group II introns and E. gracilis introns (Michel and Dujon 1983). In addition, much conserved sequences, RAGCCG.AUGAA..GAAA..UUCAUGU.CGGUUY, were found in the introns 75 nucleotides upstream from exon 2 and 61 nucleotides upstream from exon 3. This consensus sequence has been present in all the introns found so far in chloroplast genes of E. gracilis (Keller and Michel 1985). The secondary structures of introns

A

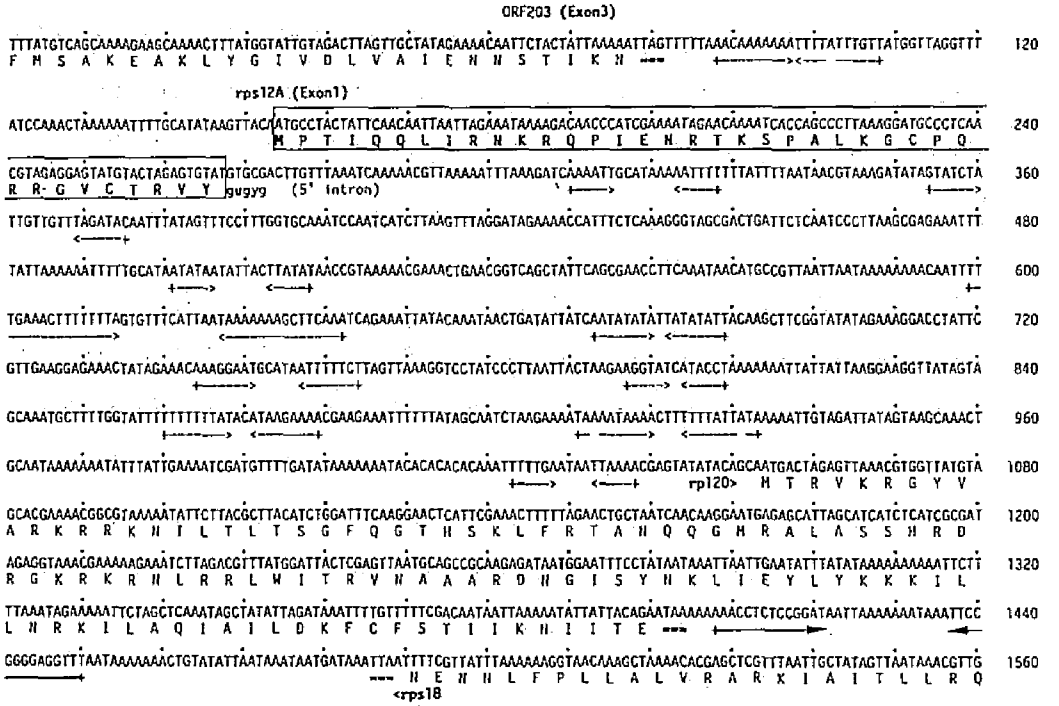


Figure 2. Complete nucleotide sequences near exon 1 (A), and exons 2 and 3 (B) of the rps12 gene. The exons of rps12 gene are boxed, and the consensus sequences of their 5' intron boundary regions (gugyg, y represents C or T residues), as well as those of the near 3' intron boundary regions (ragccg-augaa--gaaa--uucaugu-cgguuy, and cuayy(-)y-ay, r represents A or G residues) are shown under the nucleotide sequences. J_{LA} stands for a junction of an inverted repeat (IR_A) and a large single copy region (LSC). Amino acids are expressed as one letter symbols under the nucleotide sequences. Possible stem structure for the transcription termination are shown by bold arrows.

B

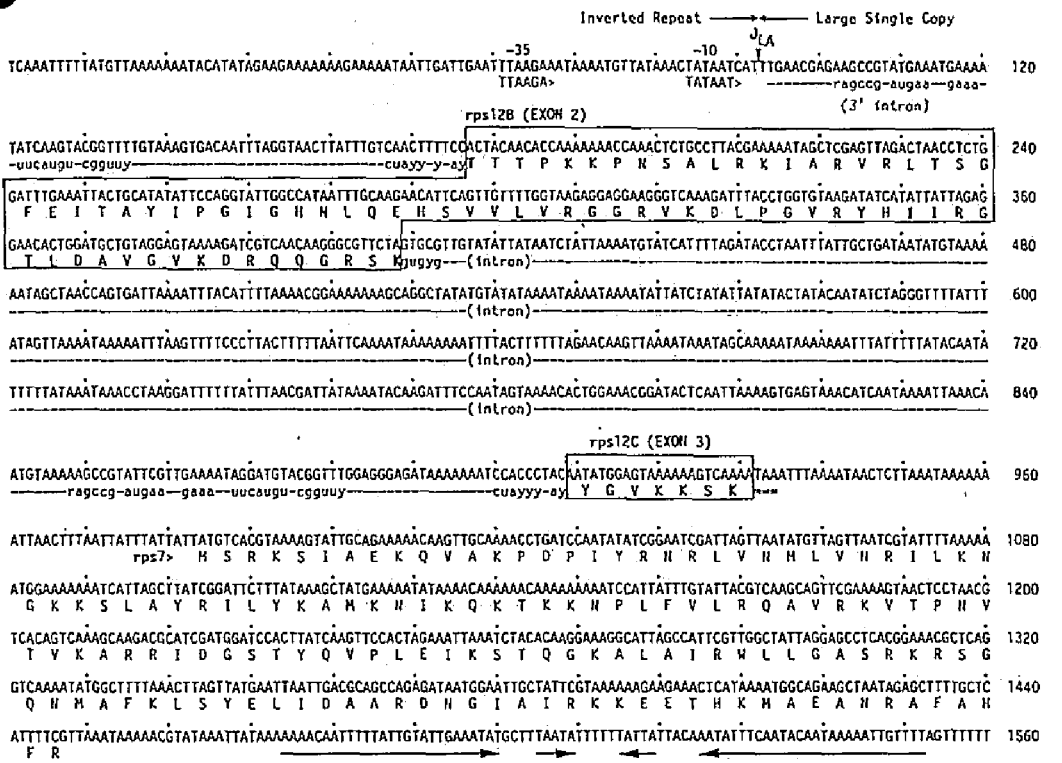


Figure 2 (continued).

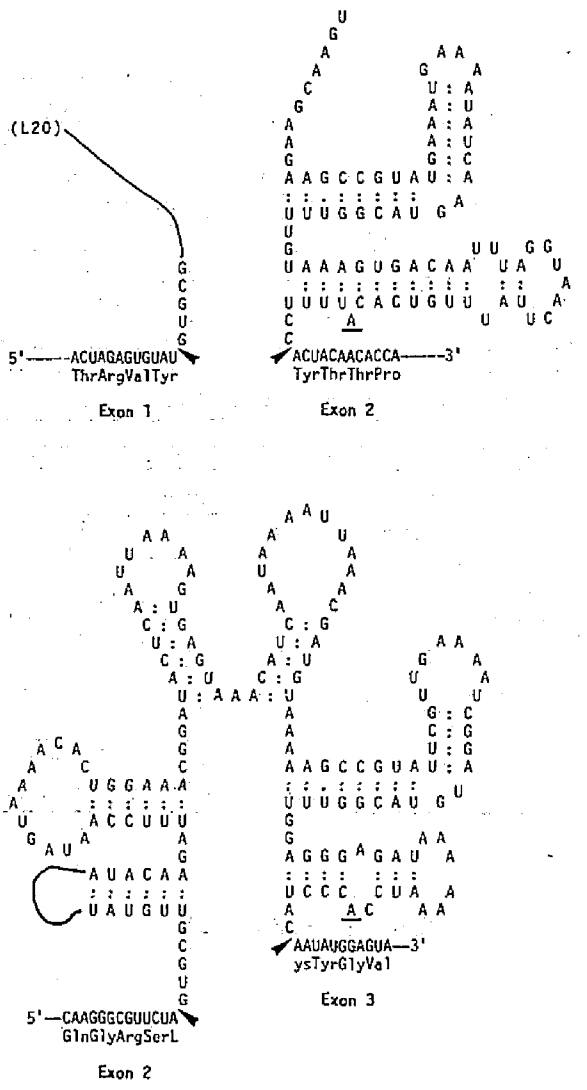


Figure 3. Possibly secondary structures of introns in rps12 gene. Intron-exon boundaries are shown by arrow heads. Two distinct RNA molecules containing exon1 and exon2 are placed close together according to group II introns specific stem-loop structure. Amino acid sequence are shown under the RNA sequences by three letter codes.

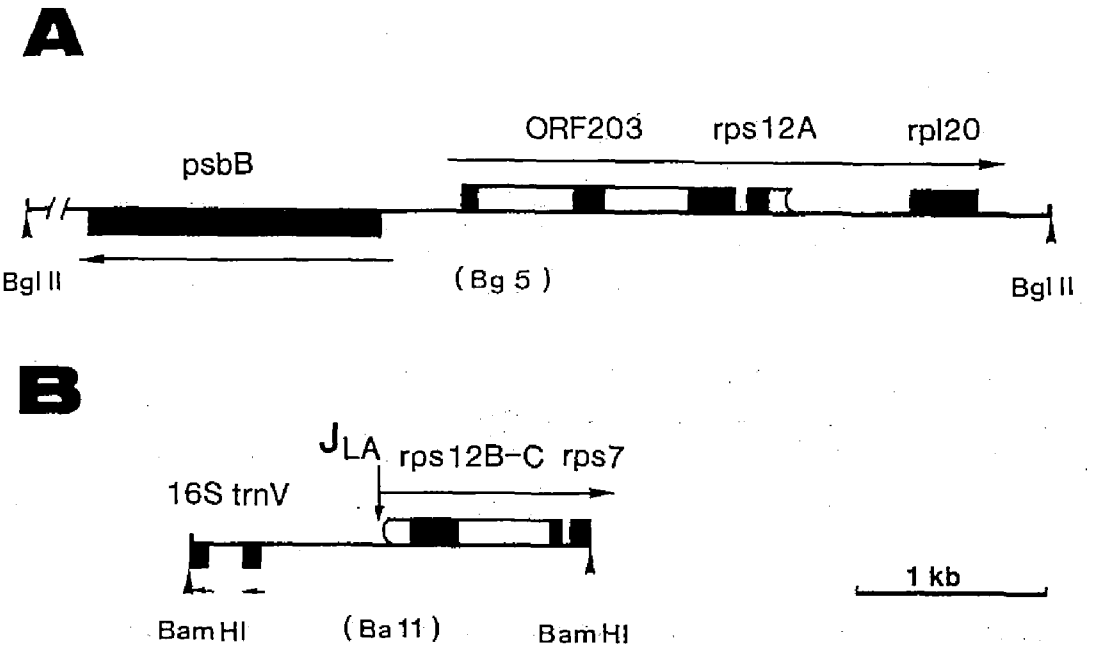


Figure 4. Gene organizations near the rps12 gene in exon 1 (A) and exons 2 and 3 (B). Arrows indicate the direction of transcription. The abbreviations rps12, rps7 and rpl20 are the same as in Fig. 1. The symbol, psbB, represents the gene for the P680 chlorophyll a apoprotein in photosystem II. J_{LA} stands for a junction of an inverted repeat (IR_A) and a large single copy region (LSC).

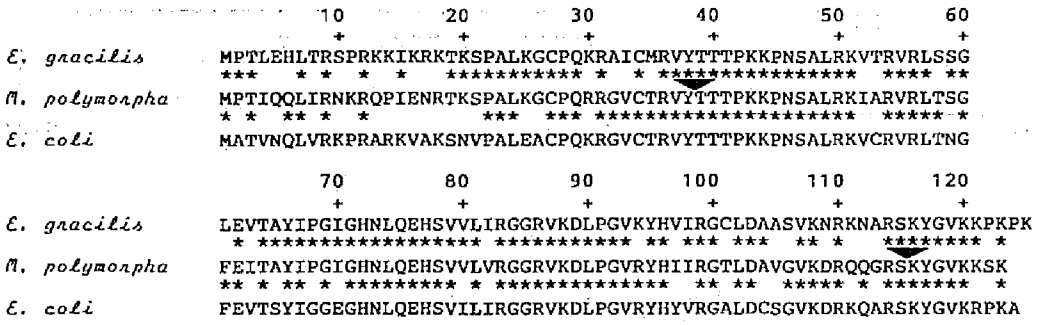


Figure 5. Amino acid sequences for the ribosomal protein S12 from M. polymorpha compared with those from E. coli and E. gracilis. Vertical arrow heads indicate sites of splicing junctions in the ribosomal protein S12 from M. polymorpha. Asterisks denote amino acids that are identical between the two proteins. Amino acids are expressed as one letter symbols.

between exons are shown in Fig. 3. Splice junctions are indicated by arrow heads deduced from the conserved sequences and secondary structures. The putative branching nucleotide (adenine) is shown by underline in Fig. 3.

The gene organizations deduced from DNA sequences near coding regions for ribosomal protein S12 are shown in Fig. 4. ORFs corresponding to the ribosomal proteins S7 and L20 were identified by their amino acid sequence homologies, 42.6% and 46.6%, for the respective E. coli ribosomal proteins (Post and Nomura 1980, Wittmann and Seib 1979). In case of the ribosomal protein S14 of M. polymorpha amino acid homology was 45% to that of E. coli (Umesono et al. 1984). By contrast, the amino acid sequence of chloroplast ribosomal protein S12 from M. polymorpha showed markedly higher homologies, 73.6% to that of E. gracilis (Montandon and Stutz 1984) and 70.2% to that of E. coli (Post et al. 1978) (Fig. 5). The amino acid sequence of the ribosomal protein S12 from M. polymorpha near the splice junctions (arrow heads in Fig. 5) showed an even higher homology to sequences from E. gracilis and E. coli, both of which have no intron. This highly conserved amino acid sequence suggests that the chloroplast ribosomal protein S12 may have an essential role in the ribosomal function during protein synthesis in chloroplasts and that rps12 gene may have a regulating function in the coordinative biogenesis of chloroplast ribosome.

The ORF203, which was interrupted by two introns (518 bp and 380 bp), was detected further upstream from exon 1 of rps12 gene (Fig. 3A). An reading frame from the ATG codon (67372) to the TAG stop codon (67130) in the third exon of ORF203 was previously called ORF80 because of assuming no intron. Although the exon-intron boundary sequences of ORF203 are a little diverged from the consensus sequences, introns are correctly excised from the precursor messenger RNA including rps12A and rpl20 coding sequences (Kohchi and Umesono unpublished data). In addition, first exon of ORF203 (68640-68570) shows 20 out of 23 amino acid residues identity (87.0%) with an reading frame in spinach X-gene next to psbB gene, which is shown to be transcribed in vivo.

Exon 1 of ribosomal protein S12 with the 5' intron boundary sequence was followed by a coding sequence of ribosomal protein L20 (Fig. 2A) 786 bp apart. Following the coding region of ribosomal protein S12 in exon 3 is a ribosomal protein S7 coding region (Fig. 2B). Close linkage of ribosomal protein S7 and S12 genes also exists in E. gracilis (Montandon and Stutz 1983) and E. coli str-operon (Post et al. 1978, Post and Nomura 1980).

Transcription for exons 2 and 3, as well as for the ribosomal protein S7 gene, is initiated by a typical prokaryotic promoter sequence (-35 and -10 regions) found upstream (Fig. 2B). S1 mappings showed that this promoter was highly active in chloroplasts as well as in E. coli (Fukuzawa et al. 1985). Northern hybridizations also showed the active transcription for exon 1. If the split gene described here provides active mRNA that could be translated into mature S12 protein, there must be a rejoining of exons at the RNA or DNA level. The DNA rearrangement of these coding regions is not observed in chloroplast genome by restriction analysis. Results of S1 mappings and Northern hybridizations suggest that transcription units for exons 2 and 3 are independent of the unit for exon 1. Therefore, active mRNA for ribosomal protein S12 was thought to be formed post-transcriptionally by a mechanism such as that of trans-splicing described by Solnick (1985) and Konarsk et al. (1985). An further investigation of the transcription and splicing mechanisms for the split gene rps12 would clarify this complex gene organization and expression in the chloroplast.

After the publication of these results, the nucleotide sequences of the tobacco chloroplast ribosomal protein S12 was reported to be trans-split as in the case of liverwort (Torazawa et al. 1986). This suggest that trans-split rps12 gene on the chloroplast genome may be a general feature in plant chloroplasts and has some regulatory function in chloroplast biogenesis.

REFERENCES

Aiba H, Adhya S, Crombrugghe B (1981) Evidence for two functional gal promoters in intact Escherichia coli cells. J Biol Chem 256:11905-11910

Aldrich J, Cherney B, Merlin E, Williams C, Mets L (1985) Recombination within the inverted repeat sequences of the Chlamydomonas reinhardii chloroplast genome produces two orientation isomers. Curr Genet 9:233-238

Alt J, Westhoff P, Sears BB, Nelson N, Hurt E, Hauska G, Herrmann RG (1983) Genes and transcripts for the polypeptides of the cytochrome b-f complex from spinach thylakoid membranes. EMBO J 2:979-986

Alt J, Winter P, Sebald W, Moser JG, Schedel R, Westhoff P, Herrmann RG (1983) Localization and nucleotide sequence of the gene for the ATP synthase proteolipid subunit on the spinach plastid chromosome. Curr Genet 7:129-138

Alt J, Herrmann RG (1984) Nucleotide sequence of the gene for pre-apocytochrome f in the spinach plastid chromosome. Curr Genet 8:551-557

Alt J, Morris J, Westhoff P, Herrmann RG (1984) Nucleotide sequence of the clustered genes for the 44 kd chlorophyll a apoprotein and the "32 kd"-like protein of the photosystem II reaction center in the spinach plastid chromosome. Curr Genet 8:597-606

Anderson S, Bankier AT, Barrell BG, Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Shreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457-465

Bartsch M, Kimura M, Subramanian AR (1982) Purification, primary structure, and homology relationship of a chloroplast ribosomal protein. Proc Natl Acad Sci USA 79:6871-6875

Bartsch M (1985) Correlation of chloroplast and bacterial ribosomal proteins by cross-reactions of antibodies specific to purified Escherichia coli ribosomal proteins. J Biol Chem 260:237-241

Bedbrook JR, Kolodner R, Bogorad L (1976) Endonuclease recognition sites mapped on Zea mays chloroplast DNA. Proc Natl Acad Sci USA 73:4309-4312

Bedwell D, Davis G, Gosink M, Post L, Nomura M, Kestler H, Zengel JM, Lindahl L (1985) Nucleotide sequence of the alpha ribosomal protein operon of Escherichia coli Nucl Acids Res 13:3891-3903

Bennoun P (1982) Evidence for a respiratory chain in the chloroplast. Proc Natl Acad Sci 79:4352-4356

Biggin MD, Gibson TJ, Hong GF (1983) Buffer gradient gels and ^{35}S label as an

aid to rapid DNA sequence determination. Proc Natl Acad Sci USA 80:3963-3965

Bird CR, Koller B, Auffret AD, Huttly AK, Howe CJ, Dyer TA, Gray JC (1985) The wheat chloroplast gene for CF_0 subunit I of ATP synthase contains a large intron. EMBO J 4:1381-1388

Bonnard G, Michel F, Weil JH, Steinmetz A (1984) Nucleotide sequence of the split $tRNA^{Leu}$(UAA) gene from Vicia faba chloroplasts: evidence for structural homologies of the chloroplast $tRNA^{Leu}$ intron with the intron from the autosplicable Tetrahymena ribosomal RNA precursor. Mol Gen Genet 194:330-336

Briat JF, Mache R (1980) Properties and characterization of a spinach chloroplast RNA polymerase isolated from a transcriptionally active DNA-protein. Eur J Biochem 111:503

Calaghan JL, Pirtle RM, Pirtle IL, Kashdan MA, Vreman, HJ, Dudock BS (1980) Homology between chloroplast and prokaryotic initiator tRNA, nucleotide sequence of spinach chloroplast methionine initiator tRNA. J Biol Chem 255:9981-9984

Canaday J, Guillemaut P, Gloeckler R, Weil JH (1981) The nucleotide sequence of spinach chloroplast tryptophan transfer RNA. Nucl Acids Res 9:47-53

Casadaban MJ, Chou J, Cohen SN (1980) In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: Escherichia coli plasmid vectors for the detection and cloning of translational initiation signals. J Bacteriol 143:971-980

Casadaban MJ, Cohen SN (1980) Analysis of gene control signals by DNA fusion and cloning in Escherichia coli. J Mol Biol 138:179-207

Carrillo N, Seyer P, Tyagi A, Herrmann RG (1986) Cytochrome b-559 genes from Oenothera hookeri and Nicotiana tabacum shows a remarkably high degree of conservation as compared to spinach. Curr Genet 10:619-624

Cerretti DP, Dean D, Davis GR, Bedwell DM, Nomura M (1983) The spc ribosomal protein operon of Escherichia coli: sequence and cotranscription of the ribosomal protein genes and a protein export gene. Nucl Acids Res 11: 2599-2616

Chomyn A, Mariottini P, Cleeter MWJ, Ragan CI, Matsuno-Yagi A, Hatefi Y, Doolittle RF, Attardi G (1985) Six unidentified reading frames of human mitochondrial DNA encode components of the respiratory-chain NADH dehydrogenase. Nature 314:592-597

Cozens AL, Walker JE, Phillips AL, Huttly AK, Gray JC (1986) A sixth subunit of ATP synthase, an F_0 component, is encoded in the pea chloroplast genome.

EMBO J 5:217-222

Crouse E J, Schmitt J M, Bohnert H J (1985) Chloroplast and cyanobacterial genomes, genes, and RNAs: a compilation. Pant Mol Biol Rep 3:43-89

Curtis SE, Haselkorn R (1983) Isolation and sequence of the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase from the cyanobacterium Anabaena 7120. Proc Natl Acad Sci USA 80:1835-1839

Curtis SE, Haselkorn R (1984) Isolation, sequence and expression of two members of the 32 Kd thylakoid membrane protein gene family from cyanobacterium Anabaena 7120. Plant Mol Biol 3:249-258

Deininger PL (1983) Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. Anal Biochem 129:216-223

Deno H, Kato K, Shinozaki K, Sugiura M (1982) Nucleotide sequences of tobacco chloroplast genes for elongator $tRNA^{Met}$ and $tRNA^{Val}$(UAC): the $tRNA^{Val}$(UAC) gene contains a long intron. Nucl Acids Res 10:7511-7520

Deno H, Shinozaki K, Sugiura M (1983) Nucleotide sequence of tobacco chloroplast gene for the α subunit of proton-translocating ATPase. Nucl Acids Res 11:2185-2191

Deno H, Shinozaki K, Sugiura M (1984) Structure and transcription pattern of a tobacco chloroplast gene coding for subunit III of proton-translocating ATPase. Gene 32:195-201

Deno H, Sugiura M (1984) Chloroplast $tRNA^{Gly}$ gene contains a long intron in the D stem: Nucleotide sequence of tobacco chloroplast gene for $tRNA^{Gly}$(UCC) and $tRNA^{Arg}$(UCU). Proc Natl Acad Sci USA 81:405-408

Dorne AM, Lescure AM, Mache R (1984) Site of synthesis of spinach chloroplast ribosomal proteins and formation of incomplete ribosomal particles in isolated chloroplasts. Plant Mol Biol 3:83-90

Dron M, Rahire M, Rochaix JD (1982) Sequence of the chloroplast DNA region of Chlamydomonas reinhardii containing the gene of the large subunit of ribulose bisphosphate carboxylase and parts of its flanking genes. J Mol Biol 162:775-793

Ellis RJ (1981) Chloroplast proteins: synthesis, transport, and assembly. Ann Rev Plant Physiol 32:111-37

Erickson JM, Rahire M, Rochaix JD (1984) Chlamydomonas reinhardii gene for the 32000 mol. wt. protein of photosystem II contains four large introns and is located entirely within the chloroplast inverted repeat. EMBO J 3:2753-2762

Erion JL (1985) Characterization of the mRNA transcripts of the maize, ribulose-1,5-bisphosphate carboxylase, large subunit gene. Plant Mol Biol

4:169-179

Filho EJ, Hartley MR, Mache R (1981) Pea chloroplast ribosomal proteins: characterization and site of synthesis. Mol Gen Genet 184:484-488

Fish LE, Kuck U, Bogorad L (1985) Two partially homologous adjacent light-inducible maize chloroplast genes encoding polypeptides of the P700 chlorophyll a-protein complex of photosystem I. J Biol Chem 260:1413-1421

Francis MA, Dudock BS (1982) Nucleotide sequence of a spinach chloroplast isoleucine tRNA. J Biol Chem 257:11195-11198

Francis M, Kashdan M, Sprouse H, Otis L, Dudock B (1982) Nucleotide sequence of a spinach chloroplast proline tRNA. Nucl Acids Res 10:2755-2758

Fromm H, Edelman M, Koller B, Goloubinoff P, Galun E (1986) The enigma of the gene coding for ribosomal protein S12 in the chloroplasts of Nicotiana. Nucl Acids Res 14:883-898

Fukuzawa H, Uchida U, Yamano Y, Ohyama K, Komano T (1985) Molecular cloning of promoters functional in Escherichia coli from chloroplast DNA of a liverwort, Marchantia polymorpha. Agric Biol Chem 49:2725-2731

Fukuzawa H, Kohchi T, Shirai H, Ohyama K, Umesono K, Inokuchi H, Ozeki H (1986) Coding sequences for chloroplast ribosomal protein S12 from a liverwort, Marchantia polymorpha, are separated far apart on the different DNA strands. FEBS Lett 198:11-15

Gantt JS, Key JL (1986) Isolation of nuclear encoded plastid ribosomal protein cDNAs. Mol Gene Genet 202:186-193

Gatenby AA, Castleton JA, Saul MW (1981) Expression in E. coli of maize and wheat chloroplast genes for large subunit of ribulose bisphosphate carboxylase. Nature 291:117-121

Gingrich JC, Hallick RB (1985) The Euglena gracilis chloroplast ribulose-1,5-bisphosphate carboxylase gene. I. Complete DNA sequence and analysis of the nine intervening sequences. J Biol Chem 260:16156-16161

Goloubinoff P, Edelman M, Hallick RB (1984) Chloroplast-coded atrazine resistance in Solanum nigrum: psbA loci from susceptible and resistant biotypes are isogenic except for a single codon change. Nucl Acids Res 12:9489-9496

Graves MC, Spremulli LL (1983) Activity of Euglena gracilis chloroplast ribosomes with procaryotic and eucariotic initiation factors. Arch Biochem Biophys 222:192

Greenberg BM, Narita JO, Flaherty CD, Gruissem W, Rushlow KA, Hallick RB (1984) Evidence for two RNA polymerase activities in Euglena gracilis chloroplasts.

J Biol Chem 259:14880

Gruissem W, Greenberg BM, Zurawski G, Prescott D, Hallick RB (1983) Biosynthesis of chloroplast transfer RNA in a spinach chloroplast transcription system. Cell 35:815-828

Gruissem W, Zurawski G (1985) Analysis of promoter regions for the spinach chloroplast rbcL, atpB and psbA genes. EMBO J 4:3375-3383

Gruissem W, Zurawski G (1985) Identification and mutational analysis of the promoter for a spinach chloroplast transfer RNA gene. EMBO J 4:1637-1644

Hallick RB, Hollingseorth MJ, Nickoloff JA (1984) Transfer RNA genes of Euglena gracilis chloroplast DNA. A revew. Plant Mol Biol 3:169-175

Hallick RB, Gingrich JC, Johanningmeier U, Passavant CW (1985) Introns in Euglena and Nicotiana chloroplast protein genes. in: Molecular form and function of the plant genome. Plenum, New York, pp211-220

Hauska G (1985) Organization and function of cytochrome b_6f/bc_1 complexes: in Molecular biology of the photosynthetic apparatus. Cold Spring Harbor Lab. New York

Heinemeyer W, Alt J, Herrmann RG (1984) Nucleotide sequence of the clustered genes for apocytochrome b6 and subunit 4 of the cytochrome b/f complex in the spinach plastid chromosome. Curr Genet 8:543-549

Hellmund D, Metzlaff M, Serfling E (1984) A transfer RNA^{Arg} gene of Pelargonium chloroplasts, but not a 5S RNA gene, is efficiently transcribed after injection into Xenopus oocyte nuclei. Nuc Acids Res 12:8253-8268

Hennig J, Herrmann RG (1986) Chloroplast ATP synthase of spinach contains nine nonidentical subunit species, six of which are encoded by plastid chromosomes in two operons in a phylogenetically conserved arrangement. Mol Gen Genet 203:117-128

Herrmann RG, Alt J, Schiller B, Widger WR, Cramer WA (1984) Nucleotide sequence of the gene for apocytochrome b-559 on the spinach plastid chromosome: implications for the structure of the membrane protein. FEBS lett 176:239-244

Higgins CF, Hiles ID, Whalley K, Jamieson DJ (1985) Nucleotide binding by membrane components of bacterial periplasmic binding protein-dependent transport systems. EMBO J 4:1033-1040

Hird SM, Willey DL, Dyer TA, Gray JC (1986) Location and nucleotide sequence of the gene for cytochrome b-559 in wheat chloroplast DNA. Mol Gen Genet 203:95-100

Hirschberg J, McIntosh L (1983) Molecular basis of herbicide resistance in Amaranthus hybridus. Science 222:1346-1349

Holshuh K, Bottomley W, Whitfeld PR (1984) Structure of the spinach chloroplast genes for the D2 and 44kd reaction-centre proteins of photosystem II and $tRNA^{Ser}(UGA)$. Nucl Acids Res 12:8819-8834

Howe CJ, Auffret AD, Doherty A, Bowman CM, Dyer TA, Gray JC (1982) Location and nucleotide sequence of the gene for the proton-translocating subunit of wheat chloroplast ATP synthase. Proc Natl Acad Sci USA 79:6903-6907

Hu N, Messing J The making of strand-specific probes. Gene 17:271-277

Karabin GD, Farley M, Hallick RB (1984) Chloroplast gene for M_R 32000 polypeptide of photosystem II in Euglena gracilis is interrupted by four introns with conserved boundary sequences. Nucl Acids Res 12:5801-5812

Kashdan MA, Dudock B (1982) The gene for a spinach chloroplast isoleucine tRNA has a methionine anticodon. J Biol Chem 257:11191-11194

Kato A, Takaiwa F, Shinozaki K, Sugiura M (1985) Location and nucleotide sequence of the gene for tobacco chloroplast $tRNA^{Arg}(ACG)$ and $tRNA^{Leu}(UAG)$. Curr Genet 9:405-409

Keller M, Michel F (1985) The introns of the Euglena gracilis chloroplast gene which codes for the 32-kDa protein of photosystem II. Evidence for structural homologies with class II introns. FEBS lett 179:69-73

Keller M, Stutz E (1984) Structure of the Euglena gracilis chloroplast gene (psbA) coding for the 32 Kd protein of photosystem II. FEBS Lett 175:173-177

Keus RJ, Stam NJ, Zwiers T, Heij HT, Groot GSP (1984) The nucleotide sequence of the genes for $tRNA^{Arg}UCU$, $tRNA^{Arg}ACG$ and $tRNA^{Asn}GUU$ on Spirodela oligorhiza chloroplast DNA. Nucl Acids Res 12:5639-5646

Kidd GH, Bogorad L (1979) Peptide maps comparing subunits of maize chloroplast and type II nuclear DNA-dependent RNA polymerases. Proc Natl Acad Sci USA 76:4890-4892

Kirsch W, Seyer P, Herrmann RG (1986) Nucleotide sequence of the cluster genes for two P_{700} chlorophyll a apoproteins of the photosystem I reaction center and the ribosomal protein S14 of the spinach plastid chromosome. Curr Genet 10:843-855

Koch W, Edwards K, Kossel H (1981) Sequencing of the 16S-23S spacer in a ribosomal RNA operon of Zea mays chloroplast DNA reveals two split tRNA genes. Cell 25:203-213

Kohchi T, Shirai H, Fukuzawa H, Sano T, Sano S, Ohyama K, Umesono K, Ozeki H (1986) Structure and organization of Marchantia polymorpha chloroplast genome. (IV) Inverted repeat and small single copy region including frx and ndh genes. In preparation.

Koller B, Gingrich JC, Stiegler GL, Farley MA, Delius H, Hallick RB (1984) Nine introns with conserved boundary sequences in the Euglena gracilis chloroplast ribulose1,5-bisphosphate carboxylase gene. Cell 36:545-553

Konarska MM, Padgett RA, Sharp PA (1985) Trans splicing of mRNA precursors in vitro. Cell 42:165-171

Kong XF, Lovett PS, Kung SD (1984) The Nicotiana chloroplast genome. IX. Identification of regions active as procaryotic promoters in Escherichia coli Gene 31:23-30

Kozak M (1983) Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol Rev 47:1-45

Krebbers ET, Larrinua M, McIntosh L, Bogorad L (1982) The maize chloroplast genes for the β and ϵ subunits of the photosynthetic coupling factor CF_1 are fused. Nucl Acids Res 10:4985-5002

Krebbers E, Steinmetz A, Bogorad L (1984) DNA sequences for the Zea mays tRNA genes tV-UAC and tS-UGA: tV-UAC contains a large intron. Plant Mol Biol 3:13-20

Langridge P (1981) Synthesis of the large subunit of spinach ribulose bisphosphate carboxylase may involve a precursor polypeptide. FEBS lett 123:85-89

Lerbs S, Brautigam E, Parthier B (1985) Polypeptides of DNA-dependent RNA polymerase of spinach chloroplasts: characterization by antibody-linked polymerase assay and determination of sites of synthesis. EMBO J 4:1661-1666

Link G, Langridge U (1984) Structure of the chloroplast gene for the precursor of the Mr 32,000 photosystem II protein from mustard (Snapis alba L.) Nucl Acids Res 12:945-958

Maniatis T, Fritsch EF, Sambrook J (1982) "Molecular cloning: A laboratory manual" Cold Spring Harbor Lab, New York

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci USA 74:560-564

McIntosh L, Poulsen C, Bogorad L (1980) Chloroplast gene sequence for the large subunit of ribulose bisphosphatecarboxylase of maize. Nature 288:556-560

Meek DW, Hayward R (1984) Nucleotide sequence of the rpoA-rplQ DNA of Escherichia coli: a second regulatory binding site for protein S4? Nucl Acids Res 12:5813-5821

Messing J (1983) New M13 Vectors for cloning. in "Methods in Enzymology", Academic Press, New York, 101: pp20-78

Mets LJ, Geist L (1983) Linkage of a known chloroplast gene mutation to the

uniparental genome of Chlamydomonas reinhardii. Genetics 105:559-579

Michel F, Dujon B (1983) Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. EMBO J 2:33-38

Montandon PE, Stutz E (1983) Nucleotide sequence of a Euglena gracilis chloroplast genome region coding for the elongation factor Tu; evidence for a spliced mRNA. Nucl Acids Res 11:5877-5892

Montandon PE, Stutz E (1984) The genes for the ribosomal protein S12 and S7 are clustered with the gene for the EF-Tu protein on the chloroplast genome of Euglena gracilis. Nucl Acid Res 12:2851-2859

Morris J, Herrmann RG (1984) Nucleotide sequence of the gene for the P_{680} chlorophyll a apoprotein of the photosystem II reaction center from spinach. Nucl Acids Res 12:2837-2850

Muller GS, Hallick RB, Alt J, Westhoff P, Herrmann RG (1986) Spinach plastid genes coding for initiation factor IR-1, ribosomal protein S11 and RNA polymerase α-subunit. Nucl Acid Res 14:1029-1044

Mullet JE, Orozco EM, Chua NH (1985) Multiple transcripts for higher plant rbcL and atpB genes and localization of the transcription initiation site of the rbcL gene. Plant Mol Biol 4:39-54

Nargang F, McIntosh L, Somerville C (1984) Nucleotide sequence of the ribulosebisphosphate carboxylase gene from Rodospirillum rubrum. Mol Gen Genet 193:220-224

Ohme M, Tanaka M, Chunwonfse J, Shinozaki K, Sugiura M (1986) A tobacco chloroplast DNA sequence possibly coding for a polypeptide similar to E. coli RNA polymerase β-subunit. FEBS lett 200:87-90

Ohyama K, Wetter LR, Yamano Y, Fukuzawa H, Komano T (1982) A simple method for isolation of chloroplast DNA from Marchantia polymorpha L. cell suspension cultures. Agric Biol Chem 46:237-242

Ohyama K, Yamano Y, Fukuzawa H, Komano T, Yamagishi H, Fujimoto S, Sugiura M (1983) Physical mappings of chloroplast DNA from a liverwort Marchantia polymorpha L. cell suspension cultures. Mol Gen Genet 189:1-9

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki S, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA. Nature 322:572-574

Oishi KK, Shapiro DR, Tewari KK (1984) Sequence organization of a pea chloroplast DNA gene coding for a 34,500-dalton protein. Mol Cell Biol

4:2556-2563

Palmer JD (1983) Chloroplast DNA exists in two orientations. Nature 301:92-93

Palmer JD (1985) Comparative organization of chloroplast genome. Ann Rev Genet 19:325-354

Perron CY, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103-119

Phillips AL, Gray JC (1984) Location and nucleotide sequence of the gene for the 15.2 kDa polypeptide of the cytochrome b-f complex from pea chloroplast. Mol Gen Genet 194:477-484

Pirtle R, Calagan J, Pirtle I, Kashdan M, Vreman H, Dudock B (1982) The nucleotide sequence of spinach chloroplast methionine elongator tRNA. Nucl Acids Res 9:183-188

Platt T, Muller-Hill B, Miller JH (1972) Assays of the lac operon enzymes. in: Experiments in molecular genetics. Cold Spring Harbor Lab., New York, pp.352

Pon CL, Wittmann LB, Gualerzi C (1979) Structure-function relationships in Escherichia coli initiation factors. II Elucidation of the primary structure of initiation factor IF-1. FEBS Lett 101:157-160

Posno M, Vliet A, Groot GSP (1986) The gene for Spirodela oligorhiza chloroplast ribosomal protein homologous to E. coli ribosomal protein L16 is split by a large intron near its 5' end: structure and expression. Nucl Acid Res 14:3181-3195

Post LE, Arfsten AE, Reusser F, Nomura M (1978) DNA sequence of promoter regions for the str and spc ribosomal protein operons in E. coli. Cell 15:215-229

Post LE, Nomura M (1980) DNA sequences from the str-operon of Escherichia coli. J Biol Chem 255:4660-4666

Pribnow D (1975) Bacteriophage T7 early promoters: Nucleotide sequences of two RNA polymerase binding sites. J Mol Biol 99:419-443

Projan SJ, Carleton S, Novick RP (1983) Determination of plasmid copy number by fluorescence densitometry. Plasmid 9:182-190

Quigley F, Weil JH (1985) Organization and sequence of five tRNA genes and of an unidentified reading frame in the wheat chloroplast genome: evidence for gene rearrangements during the evolution of chloroplast genomes. Curr Genet 9:495-503

Rasmussen OF, Bookjans G, Stummann BM, Henningsen KW (1984) Localization and nucleotide sequence of the gene for the membrane polypeptide D2 from pea chloroplast DNA. Plant Mol Biol 3:191-199

Reichelt BY, Delaney SF (1983) The nucleotide sequence for the large subunit of ribulose 1,5-bisphosphate carboxylase from a unicellular cyanobacterium, Synechococcus PCC6301. DNA 2:121-129

Rochaix JD, Malnoe P (1978) Anatomy of the chloroplast ribosomal DNA of Chlamydomonas reinhardii. Cell 15:661-670

Rochaix JD, Dron M, Rahire M, Malnoe P (1985) Sequence homology between the 32K dalton and the D2 chloroplast membrane polypeptides of Chlamydomonas reinhardii. Plant Mol Biol 3:363-370

Rochaix JD, Rahire M, Michel F (1985) The chloroplast ribosomal intron of Chlamydomonas reinhardii codes for a polypeptide related to mitochondrial maturase. Nucl Acids Res 13:975-984

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74:5463-5467

Saraste M, Gay NJ, Eberle A, Runswick MJ, Walker J (1981) The atp operon: nucleotide sequence of the genes for the r, b, and e subunits of Escherichia coli ATP synthase. Nucl Acids Res 9:5287-5296

Schmidt RJ, Richardson CB, Gillham NW, Boynton JE (1983) Sites of synthesis of chloroplast ribosomal proteins in Chlamydomonas. J Cell Biol 96:1451-1463

Schmidt RJ, Hosler JP, Gillham NW, Boynton JE (1985) Biogenesis and evolution of chloroplast ribosomes: cooperation of nuclear and chloroplast genes. In: Molecular biology of the photosynthetic apparatus. Cold Spring Harbor Lab, New York, pp417-427

Schwarz Zs, Kossel H (1980) The primary structure of 16S rDNA from Zea mays chloroplast is homologous to E. coli 16S rRNA. Nature 283:739-742

Shine L, Dalgarno L (1974) The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: Complementary to nonsense triplets and ribosome binding sites. Proc Natl Acad Sci USA 71:1342-1346

Shinozaki K, Sugiura M (1982) The nucleotide sequence of the tobacco chloroplast gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. Gene 20:91-102

Shinozaki K, Deno H, Kato A, Sugiura M (1983) Overlap and cotranscription of the genes for the beta and epsilon subunits of tobacco chloroplast ATPase. Gene 24:147-155

Shinozaki K, Yamada C, Takahata N, Sugiura M (1983) Molecular cloning and sequence analysis of the cyanobacterial gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. Proc Natl Acad Sci USA 80:4050-4054

Shinozaki K, Deno H, Wakasugi T, Sugiura M (1986) Tobacco chloroplast gene coding for subunit I of proton-translocating ATPase: comparison with the wheat subunit I and E. coli subunit b. Curr Genet 10:421-423

Shinozaki K, Deno H, Sugita M, Kuramotsu S, Sugiura M (1986) Intron in the gene for the ribosomal protein S16 of tobacco chloroplast and its conserved boundary sequences. Mol Gen Genet 202:1-5

Solnick D (1985) Trans splicing of mRNA precursors. Cell 42:157-164

Spielmann A, Stutz E (1983) Nucleotide sequence of soybean chloroplast DNA regions which contain the psbA and trnH genes and cover the ends of the large single copy region and one end of the inverted repeats. Nucl Acids Res 11:7157-7167

Steinmetz AA, Gubbins EJ, Bogorad L (1983) The anticodon of the maize chloroplast gene for $tRNA^{Leu}_{UAA}$ is split by a large intron. Nucl Acids Res 10:3027-3037

Steinmetz AA, Castroviejo M, Sayre RT, Bogorad L (1986) Protein PSII-G. An additional component of photosystem II identified through its plastid gene in maize. J Biol Chem 261:2485-2488

Strittmatter G, Gozdzicka-Jozefiak A, Kossel H (1985) Identification of an rRNA operon promoter from Zea mays chloroplasts which excludes the proximal $tRNA^{Val}_{GAC}$ from the primary transcript. EMBO J 4:599-604

Subramanian AR, Steinmetz A, Bogorad L (1983) Maize chloroplast DNA encodes a protein homologous to the bacterial ribosomal assembly protein S4. Nucl Acids Res 11:5211-5286

Sugita M, Sugiura M (1983) A putative gene for tobacco chloroplast coding for ribosomal protein similar to E. coli ribosomal protein S19. Nucl Acids Res 11:1913-1918

Sugita M, Sugiura M (1984) Nucleotide sequence and transcription of the gene for the 32,000 dalton thylakoid membrane protein from Nicotiana tabacum. Mol Gen Genet 195:308-313

Sugita M, Shinozaki K, Sugiura M (1985) Tobacco chloroplast $tRNA^{Lys}$(UUU) gene contains a 2.5-kilobasepair intron: an open reading frame and a conserved boundary sequence in the intron. Proc Natl Acad Sci USA 82:3557-3561

Takaiwa F, Sugiura M (1982) Nucleotide sequence of the 16S-23S spacer region in an rRNA gene cluster from tobacco chloroplast DNA. Nucl Acids Res 10:2665-2676

Takanami M, Sugimoto K, Sugisaki H, Okamoto T (1976) Sequence of promoter for coat protein gene of bacteriophage fd. Nature 260:297-302

Tewari KK, Goel A (1983) Solubilization and partial purification of RNA polymerase from pea chloroplasts. Biochem 22:2142-2148

Tohdoh N, Sugiura M (1982) The complete nucleotide sequence of a 16S ribosomal RNA gene from tobacco chloroplasts. Gene 17:213-218

Torazawa K, Hayashida N, Obokata J, Shinozaki K, Sugiura M (1986) The 5' part of the gene for ribosomal protein S12 is located 30 kbp downstream from its 3' part in tobacco chloroplast genome. Nucl Acids Res 14:3143

Tyagi AK, Herrmann RG (1986) Location and nucleotide sequence of the pre-apocytochrome f gene on the Oenothera hookeri plastid chromosome (Euoenothera plastome I). Curr Genet 10:481-486

Umesono K, Inokuchi H, Ohyama K, Ozeki H (1984) Nucleotide sequence of Marchantia polymorpha chloroplast DNA: a region possibly encoding three tRNAs and three proteins including a homologue of E. coli ribosomal protein S14. Nucl Acids Res 12:9551-9565

Umesono K, Inokuchi H, Shiki Y, Takeuchi M, Chang Z, Fukuzawa H, Kohchi T, Sano T, Ohyama K, Ozeki H (1986) Structure and organization of Marchantia polymorpha chloroplast genome. (II) Inverted orientation of a 25 kbp portion in LSC region. in preparation.

Westhoff P (1985) Transcription of the gene encoding the 51 kd chlorophyll a-apoprotein of the photosystem II reaction centre from spinach. Mol Gen Genet 201:115-123

Whitfeld PR, Bottomley W (1983) Organization and structure of chloroplast genes. Ann Rev Plant Physiol 34:279-310

Widger WR, Cramer WA, Herrmann RG, Trebst A (1984) Sequence homology and structural similarity between cytochrome b of mitochondrial complex III and the chloroplast b_6-f complex: position of the cytochrome b hemes in the membrane. Proc Natl Acad Sci USA 81:674-678

Wilbur WJ, Lipman DJ (1983) Rapid similarity searches of nucleic acid and protein data banks. Proc Natl Acad Sci USA 80:726-730

Willey DL, Huttly AK, Phillips AL, Gray JC (1984) Localization of the gene for cytochrome f in pea chloroplast DNA. Mol Gen Genet 189:85-89

Willey DL, Auffret AD, Gray JC (1984) Structure and topology of cytochrome f in pea chloroplast membranes. Cell 36:555-562

Willey DL, Howe CJ, Auffret AD, Bowman CM, Dyer TA, Gray JC (1984) Location and nucleotide sequence of the gene for cytochrome f in wheat chloroplast DNA. Mol Gen Genet 194:416-422

Wittmann-Liebold B, Seib C (1979) The primary structure of protein L20 from the

large subunit of the Escherichia coli ribosome. FEBS lett 103:61-65

Yamano Y, Ohyama K, Komano K (1984) Nucleotide sequence of chloroplast 5S ribosomal RNA from cell suspension culture of the liverwort Marchantia polymorpha and Jungermannia subulata. Nucl Acids Res 12:4621-4624

Yamano Y, Kohchi T, Fukuzawa H, Ohyama K, Komano T (1985) Nucleotide sequences of chloroplast 4.5S ribosomal RNA from a leafy liverwort, Jungermannia subulata, and a thalloid liverwort, Marchantia polymorpha. FEBS lett 185:203-207

Perron YC, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103-119

Zurawski G, Perrot B, Bottomley W, Whitfeld PR (1981) The structure of the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase from spinach chloroplast DNA. Nucl Acids Res 9:3251-3270

Zurawski G, Bottomley W, Whitfeld PR (1982) Structure of the genes for the β and ϵ subunits of spinach chloroplast ATPase indicate a dicistronic mRNA and an overlapping translation stop/start signal. Proc Natl Acad Sci USA 79:6260-6264

Zurawski G, Bohnert HJ, Whitfeld PR, Bottomley W (1982) Nucleotide sequence of the gene for the M_R 32,000 thylakoid membrane protein from Spinacia oleracea and Nicotiana debneyi predicts a totally conserved primary translation product of M_R 38,950. Proc Natl Acad Sci USA 79:7699-7703

Zurawski G, Clegg MT (1984) The Barley chloroplast DNA atpBE, trnM2, and trnV1 loci. Nucl Acids Res 12:2549-2559

Zurawski G, Clegg MT, Brown AH (1984) The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics 106:735-749

Zurawski G, Bottomley W, Whitfeld PR (1984) Junctions of the large single copy region and the inverted repeats in Spinacia oleracea and Nicotiana debneyi chloroplast DNA: sequence of the genes for $tRNA^{His}$ and the ribosomal protein S19 and L2. Nucl Acids Res 12:6547-6558

Zurawski G, Zurawski SM (1985) Structure of the Escherichia coli S10 ribosomal protein operon. Nucl Acids Res 13:4521-4526

SUMMARY

CHAPTER I Molecular cloning of promoters functional in Escherichia coli from chloroplast DNA

DNA fragments cloned from chloroplast DNA of a liverwort, Marchantia polymorpha, functional in E. coli as transcriptional promoters using gene fusion to the E. coli lac'Z gene. A recombinant plasmid gave as high a level of β-galactosidase activity as when it was induced by IPTG in E. coli wild type strain W3110. The inserted chloroplast DNA fragment was sequenced and mapped at the terminus of the inverted repeat region upstream from the 16S ribosomal RNA gene. The direction of the transcription from this promoter was opposite from that of 16S ribosomal RNA gene. This highly active promoter was for trnI-C*AU gene and clustered genes for ribosomal proteins. S1 nuclease mappings using both chloroplast and E. coli RNAs showed that the transcription starts at almost the same position downstream from the consensus Pribnow-box-like region. This clone also had a higher activity of β-galactosidase in E. coli than those containing promoters of rbcL and the β subunit gene of H^+-ATP synthase. Two clusters of genes for ribosomal proteins were identified downstream from this highly active promoters.

CHAPTER II Structure and gene organization of the chloroplast genome

The nucleotide sequence of the large single copy region (psbG-16S rRNA gene; 30,600 bp) of the chloroplast DNA from a liverwort, M. polymorpha was determined. This region encodes genes for seven tRNAs; $tRNA^{Val}$(GAC), $tRNA^{Ile}$(C*AU), $tRNA^{Arg}$(CCG), $tRNA^{Pro}$(UGG), $tRNA^{Trp}$(CCA), $tRNA^{Met}$(CAU), $tRNA^{Val}$(UAC), ten photosynthetic polypeptides; the large subunit of ribulose-1.5-bisphosphate carboxylase/oxygenase (rbcL), 51 kd photosystem II chlorophyll a apoprotein (psbB), apocytochrome b-559 polypeptides (psbE and psbF), cytochrome f preprotein (petA), cytochrome b6 polypeptide (petB) and cytochrome b6/f complex subunit 4 polypeptide (petD), β and ε subunits of H^+-ATP synthase (atpB and atpE), photosystem II G-protein (psbG), and ribosomal proteins (L2, L14, L16, L20, L22, L23, L33, S3, S8,

S11, S12, S18, S18 and S19), initiation factor 1 (infA) and α subunit of RNA polymerase (rpoA). Interestingly, functionally related genes are clustered as follow: (1) A ribosomal protein gene cluster involving transcriptional and translational machinery, trnI-L23-L2-S19-L22-S3-L16-L14-S8-infA-secX-S11-rpoA, was found at the terminus of the large single copy region next to the inverted repeat region (IR_B). (2) A cluster of photosynthetic genes, psbB-ORF35-ORF27-ORF74-petB-petD is located next to the ribosomal protein gene cluster. (3) A cluster including photosynthetic genes rbcL-trnR-ORF316-ORF36b-ORF184-ORF434-petA, was also found in large single copy region. Introns (intervening sequences) were found in coding sequences for ribosomal protein genes (rpl2, rpl16 and rps12), $tRNA^{Val}$(UAC) gene and photosynthetic genes (petB and petD). Interestingly, an open reading frame was found to show significant amino acid sequence homology to a subunit of NADH dehydrogenase in human mitochondria.

CHAPTER III Split gene for chloroplast ribosomal protein S12

A coding sequence corresponding to the E. coli ribosomal protein S12 gene (rps12) was found to be split into three exons. Strikingly, the first exon with the 5' intron boundary sequence was located on the opposite strand of the chloroplast DNA (121 kb, circular molecule) approximately 60 kb away from the rest of the exons. The amino acid sequence deduced from the DNA sequence was highly homologous to the sequences of the S12 ribosomal protein of E. coli (70.2%), and Euglena gracilis chloroplasts (73.6%). As the DNA rearrangement of these coding regions is not observed, the active messenger RNA for ribosomal protein S12 is thought to be formed post-transcriptionally such as that of trans-splicing. This may be the first identification of an example for in vivo trans-splicing.

LIST OF PUBLICATIONS

(a) Ohyama K, Wetter LR, Yamano Y, Fukuzawa H, Komano T (1982)
"A simple method for isolation of chloroplast DNA from Marchantia polymorpha L. cell suspension culture."
Agric Biol Chem 46:237-242

(b) Ohyama K, Yamano Y, Fukuzawa H, Komano T, Yamagishi H, Fujimoto S, Sugiura M (1983)
"Physical mappings of chloroplast DNA from liverwort Marchantia polymorpha L. cell suspension cultures."
Mol Gen Genet 189:1-9

(c) Tanaka A, Yamano Y, Fukuzawa H, Ohyama K, Komano T (1984)
"in vitro DNA synthesis by chloroplasts isolated from Marchantia polymorpha L. cell suspension cultures."
Agric Biol Chem 48:1239-1244

(d) Yamano Y, Kohchi T, Fukuzawa H, Ohyama K, Komano T (1985)
"Nucleotide sequence of chloroplast 4.5S ribosomal RNA from a leafy liverwort, Jungermannia subulata, and a thalloid liverwort, Marchantia polymorpha."
FEBS Letter 185:203-207

(e) Fukuzawa H, Uchida Y, Yamano Y, Ohyama K, Komano T (1985)
"Molecular cloning of promoters functional in Escherichia coli from chloroplast DNA of a liverwort Marchantia polymorpha."
Agric. Biol. Chem. (1985) 49:2725-2731

(f) Fukuzawa H, Kohchi T, Sano T, Shirai H, Umesono K, Ohyama K, Ozeki H, Komano T (1986)
"Structure and organization of Marchantia polymorpha chloroplast genome (III) LSC region (rbcL-J_{LB}) having gene clusters of ribosomal proteins and photosynthetic polypeptides."
in preparation.

(g) Fukuzawa H, Kohchi T, Shirai H, Ohyama K, Umesono K, Inokuchi H, Ozeki H (1986)
"Coding sequences for chloroplast ribosomal protein S12 from the liverwort, Marchantia polymorpha, are separated far apart on the different DNA strands."
FEBS Letter (1986) 198:11-15

(h) Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki S, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986)
"Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA."
Nature 322:572-574

Chapter I is described in reference (e).
Chapter II is described in reference (f).
Chapter III is described in reference (g).

ACKNOWLEDGMENT

The author wishes to express his grateful acknowledgment to Professor Tohru Komano, Faculty of Agriculture, Kyoto University, for his continuous encouragement throughout this work, and for critical reading of this manuscript.

The author would like to express his great gratitude to Dr. Kanji Ohyama, Laboratory of Plant Molecular Biology, Research Center for Cell and Tissue Culture, Faculty of Agriculture, Kyoto University, for allowing to use his facilities and his kind guidance throughout this study.

The author thanks to Professor Haruo Ozeki, Faculty of Science, Kyoto University, and also to Drs. Kazuhiko Umesono, Yoshiaki Yamano, Yuko Uchida, for their helpful suggestion and valuable discussion.

The author thanks to Dr. Toshimichi Ikemura, National Institute of Genetics and Dr. Minoru Kanehisa, Chemical research institute, Kyoto University, for their kind support to computer analysis using data base.

The author is indebted to Mr. Takayuki Kohchi, Mr. Tohru Sano, Mr. Hiromasa Shirai, Mr. Shigehiro Asano, Mr. Yoshinori Yamashita and all the members of the laboratory of Plant Molecular Biology, Research Center for Cell and Tissue Culture, and the laboratory of Biochemistry, Department of Agricultural Chemistry, Faculty of Agriculture, Kyoto University, for their kind assistances in this study.

Hideya Fukuzawa